Mon. Not. R. Astron. Soc. **336,** 1304–1328 (2002)

Gabor transforms on the sphere with applications to CMB power spectrum estimation

Frode K. Hansen,[1⋆] Krzysztof M. Górski[2,3⋆] and Eric Hivon[4⋆]

[1] *Dipartimento di Fisica, Università di Roma 'Tor Vergata', Via della Ricerca Scientifica 1, I-00133 Roma, Italy*
[2] *ESO, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany*
[3] *Warsaw University Observatory, Aleje Ujazdowskie 4,00-478 Warszawa, Poland*
[4] *IPAC/Caltech, Mail Code 100-22, 770 S. Wilson Av., Pasadena, CA 91125, USA*

Accepted 2002 July 17. Received 2002 July 17; in original form 2002 April 19

ABSTRACT

The Fourier transform of a data set apodized with a window function is known as the Gabor transform. In this paper we extend the Gabor transform formalism to the sphere with the intention of applying it to cosmic microwave background (CMB) data analysis. The Gabor coefficients on the sphere known as the pseudo power spectrum is studied for windows of different size. By assuming that the pseudo power spectrum coefficients are Gaussian distributed, we formulate a likelihood ansatz using these as input parameters to estimate the full-sky power spectrum from a patch on the sky. As this likelihood can be calculated quickly without having to invert huge matrices, this allows for fast power spectrum estimation. By using the pseudo power spectrum from several patches on the sky together, the full-sky power spectrum can be estimated from full-sky or nearly full-sky observations.

Key words: methods: data analysis – methods: statistical – techniques: image processing – cosmological parameters – cosmology: observations.

1 INTRODUCTION

The cosmic microwave background (CMB) is one of our most important sources of information about the early Universe (Bond 1995; Jungman et al. 1996; Hu, Sugiyama & Silk 1997; Durrer 2001). The pattern of the temperature fluctuations in the CMB contains information about a number of cosmological parameters. If the temperature fluctuations are Gaussian, as predicted by most models of the early Universe, all this information is stored in the angular power spectrum coefficients C_ℓ. For this reason, several experiments have been conducted to measure the CMB power spectrum. The *COBE* satellite discovered the fluctuations in 1992 (Smoot et al. 1992), and since then several ground-based and balloon-borne experiments (De Bernardis et al. 2000; Hanany et al. 2000; Lee et al. 2001; Halverson et al. 2001; Netterfield et al. 2002; Pryke et al. 2002) have been made to study the CMB at an ever increasing resolution. As the amount of CMB data from these experiments is rapidly growing, the task of extracting the power spectrum from the data is getting harder.

Analysing the CMB data from a given experiment consists of several steps, as the data consists of several components not belonging to the CMB (Stolyarov et al. 2001; Maino et al. 2002). In this paper, we will concentrate on extracting the power spectrum from a CMB map with foregrounds removed. The standard method of extracting the power spectrum from a sky map is the method of maximum likelihood. This method gives the smallest error bars on the power spectrum estimates, but has the drawback that the number of operations needed to perform the estimation scales as $N_{\rm pix}^3$, where $N_{\rm pix}$ is the number of pixels in the map. For experiments with high resolution, the number of pixels can be up to several million and this method becomes infeasible using current computers (Borrill 1999).

In (Oh, Spergel & Hinshaw 1999), it is shown how the likelihood analysis can be speeded up to scale as $N_{\rm pix}^2$ with assumptions about azimuthal symmetry and uncorrelated noise. Another $N_{\rm pix}^{3/2}$ method for large azimuthally symmetric parts of the sky with uncorrelated noise was presented in (Wandelt, Górski & Hivon 2001). The likelihood problem can also be solved exact in $N_{\rm pix}^2$ operations with correlated noise for special scanning strategies as demonstrated in (Wandelt 2000; Wandelt & Hansen 2001). In (Bond 1995; Bond, Jaffe & Knox 2000; Bartlett 2000) it is shown how one can approximate the likelihood to speed up the calculations, but still an $N_{\rm pix}^3$ operation is needed. This has led people to find other estimators than the maximum likelihood estimator in order to extract the power spectrum. In (Tegmark, Taylor & Heavens 1997) an optimal estimator was found but the calculation scales as $N_{\rm pix}^2$ times a huge prefactor. Recently some near-optimal estimators have been found which can be calculated in $N_{\rm pix}^2$ operations (Dore, Knox & Peel 2001; Szapudi 2001; Hivon et al. 2002) The data from the BOOMERANG (De Bernardis et al. 2000; Netterfield et al. 2002) experiment was analysed using the Monte Carlo Apodised Spherical Transform Estimator (MASTER) method (Hivon

⋆E-mail: frodekh@roma2.infn.it (FKH); kgorski@eso.org (KMG); efh@ipac.caltech.edu (EH)

et al. 2002). In this method, the power spectrum was extracted by a quadratic estimator based on the pseudo power spectrum (the power spectrum on the cut sky). A similar method was suggested by (Balbi et al. 2002) for the *Planck* surveyor. Here we propose to use the pseudo power spectrum ($\tilde{C}_\ell$) for likelihood estimation. This principle was also used in (Wandelt et al. 2001), but for large sky coverage so that the correlations between the $\tilde{C}_\ell$ coefficients could be neglected.

In this paper, we study the effect of Gabor transforms on the sphere. Gabor transforms, or windowed Fourier transforms, are just Fourier transforms where the function $f(x)$ to be Fourier transformed is multiplied with a Gabor window $W(x)$ (Gabor 1946). In the discrete case $f(x_i)$ can be a data stream. If parts of the data stream are of poor quality or are missing, this can be represented as $W(x_i)f(x_i)$ where the window W is zero where there are missing parts. The window can also be formed so that it smoothes the edges close to the missing parts and in this way avoid ringing in the Fourier spectrum.

We will study the effect of Gabor transforms on the sphere and use it for fast CMB power spectrum estimation. The Gabor transform in this context is just the multiplication of the CMB sky with a window function before using the spherical harmonic transform to get the Gabor transform coefficients in this case called the pseudo power spectrum. The window can be a top-hat to take out certain parts of the sky in the case of limited sky coverage. Another window can be a Gaussian Gabor window for smoothing the transition between the observed and unobserved area of the sky. The Gabor window can also be designed in such a way as to increase signal-to-noise ratio by giving pixels with high signal-to-noise ratios higher significance in the analysis. The use of the windowed Fourier transform was already studied in (Hobson & Magueijo 1996) in the flat-sky approximation. We show that some of their results are also valid on the sphere.

In the standard likelihood approach of power spectrum estimation, the pixels on the CMB sky or the spherical harmonic coefficients $a_{\ell m}$ are used as elements in the data vector, in which case the correlation matrix will have dimensions of the order $N_{\rm pix} \times N_{\rm pix}$. A matrix of this size can not be inverted in a reasonable amount of time with current computers. We propose to use the pseudo power spectrum coefficients $\tilde{C}_\ell$ as elements of the data vector in the likelihood. In this case the size of the correlation matrix will at most be $l_{\rm max} \times l_{\rm max}$ which can be inverted in a few seconds. The most time consuming part is the calculation of the elements of the correlation matrix of pseudo-C_ℓ.

In Section (2) we will first describe the one-dimensional Gabor transform and then define the Gabor transform on the sphere. We will define the pseudo power spectrum, which is just the Gabor coefficients on the CMB sky. The kernel relating the full-sky power spectrum and the pseudo power spectrum for a Gaussian and top-hat Gabor window will be discussed. Then in Section (3) we will use the pseudo power spectrum as input values to a maximum likelihood estimation of the full-sky power spectrum. The probability distribution of the pseudo power spectrum coefficients will be assumed Gaussian and we will show that this is a good approximation at high multipoles ($\ell > 100$). Some examples of likelihood estimations of the power spectrum with different noise patterns will be shown. In Section (4) two extensions of the method will be discussed. First the use of the pseudo power spectrum from different Gabor windows centred at different points on the sphere simultaneously is demonstrated. In this way full-sky or nearly full-sky observations can be analysed. The second extension of the method is the use of Monte Carlo simulations to obtain noise properties in the case where this is faster than using the analytic expression or where the noise is correlated. Finally in Section (5) the results and further extensions are discussed.

2 THE GABOR TRANSFORMATION AND THE TEMPERATURE POWER SPECTRUM

In this section we will first describe the Gabor transform for functions on a one-dimensional line. Then we extend the formalism to functions on the sphere and the properties of the Gabor transform coefficients on the CMB sky, the pseudo-C_ℓ, are discussed. As most CMB experiment will not be able to observe the full sky, it is important to study the properties of the power spectrum on the sky apodized with a window function. As we will show later, the best way to construct the window is not always to set it to 1 in the observed area and to 0 in the non-observed area of the sky. For this reason we will study the Gabor transform for windows with different profiles. On the cut sky the pseudo power spectrum coefficients will get coupled (Wandelt et al. 2001; Hivon et al. 2002). We will study how strong this coupling is for different window sizes and for different windows. We will in particular study the top-hat and the Gaussian windows. The top-hat window is important, as it is the window which preserves most of the information in the observed data set. The Gaussian window smoothes of the edges between the observed and unobserved areas of the sky and in this way cuts off long-range correlations between pseudo C_ℓ.

2.1 The one-dimensional Gabor transform

For a data set d_j with N elements, the normal Fourier transform is defined as,

$$\tilde{d}_k = \sum_j d_j e^{i2\pi jk/N}. \quad (1)$$

A tilde on $\tilde{d}$ shows that these are the Fourier coefficients. The inverse transform is then,

$$d_j = \frac{1}{N} \sum_k \tilde{d}_k e^{-i2\pi jk/N}. \quad (2)$$

Sometimes it is useful to study the spectrum of just a part of the data set. This could be if some parts are of poor quality or the spectrum is changing along the data set. In this case, one can multiply the data set with a function, removing the unwanted parts and taking out a segment to be studied. The function can be a step function cutting out the segment to study with sharp edges or a function which smoothes the edges of the segment to avoid ringing (typically a Gaussian).

The Fourier transform with such a multiplication was studied by Gabor (Gabor 1946) and is called the Gabor Transform. It is defined for a segment centred at $j = M$ and with wavenumber k as,

$$\tilde{d}_{kM} = \sum_j d_j G_{j-M} e^{i2\pi jk/N}. \quad (3)$$

Here G_{j-M} is the Gabor window, the function to multiply the data set with. The transform is similar to the wavelet transform. The difference is that the window function in the wavelet transform is frequency-dependent so that the size of the segment is changing with frequency.

Analogously to the Fourier transform, there is also an inverse Gabor transform. To recover the whole data set from a Gabor transform, one needs the Fourier coefficients taken with several windows G_{j-M} being centred at different points M. This means that the data set has to be split into several segments. The centre of each segment

is set to $M = mK$, where K determines the density of segments and m is an integer specifying the segment number. One then has for the inverse transform

$$d_j = \sum_m \sum_k \tilde{d}_{km} g_{km}. \tag{4}$$

Due to the non-orthogonality of the Gabor transform, the dual Gabor window g_{km} is not trivial to find, but several techniques have been developed for calculating this dual window (e.g. (Strohmer 1997) and references therein).

In this paper we will study the Gabor transform on the sphere and apply it to CMB analysis. We will take out a disc on the CMB sky, using either top-hat or Gaussian apodization and then derive the pseudo power spectrum $\tilde{C}_\ell$ on the apodized sky. The $\tilde{C}_\ell$ will be used for likelihood estimation of the underlying full-sky power spectrum. We also show how several discs (segments) centred at different points can be combined to yield the full-sky power spectrum.

2.2 Gabor transform on the sphere

We start by defining the $\tilde{C}_\ell$ for a Gabor window $G(\hat{\boldsymbol{n}})$ as,

$$\tilde{C}_\ell = \sum_m \frac{\tilde{a}^*_{\ell m}\tilde{a}_{\ell m}}{2\ell+1}, \tag{5}$$

where

$$\tilde{a}_{\ell m} = \int d\hat{\boldsymbol{n}} T(\hat{\boldsymbol{n}}) G(\hat{\boldsymbol{n}}) Y^*_{\ell m}(\hat{\boldsymbol{n}}). \tag{6}$$

Here $T(\hat{\boldsymbol{n}})$ is the observed temperature in the direction of the unit vector $\hat{\boldsymbol{n}}$, $Y_{\ell m}(\hat{\boldsymbol{n}})$ is the spherical Harmonic function and $G(\hat{\boldsymbol{n}})$ is the Gabor window. We now find an expression for the expectation value of $\tilde{C}_\ell$.

Here we will use a Gabor window which is azimuthally symmetric about a point $\hat{\boldsymbol{n}}_0$ on the sphere, so that the window is only a function of the angular distance from this point on the sphere $\cos\theta = \hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{n}}_0$. Then one can write the Legendre expansion of the window as,

$$G(\theta) = \sum_\ell \frac{2\ell+1}{4\pi} g_\ell P_\ell(\cos\theta) = \sum_{\ell m} g_\ell Y_{\ell m}(\hat{\boldsymbol{n}}) Y^*_{\ell m}(\hat{\boldsymbol{n}}_0). \tag{7}$$

One can also write,

$$T(\hat{\boldsymbol{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}}). \tag{8}$$

Inserting these two expressions in equation (6) one gets

$$\begin{aligned}\tilde{a}_{\ell m} &= \sum_{\ell' m'} a_{\ell' m'} \sum_{\ell'' m''} g_{\ell''} Y^*_{\ell'' m''}(\boldsymbol{n}_0) \int Y^*_{\ell m}(\hat{\boldsymbol{n}}) Y_{\ell' m'}(\hat{\boldsymbol{n}}) Y_{\ell'' m''}(\hat{\boldsymbol{n}})\, d\hat{\boldsymbol{n}} \\ &= \sum_{\ell' m'} a_{\ell' m'} \sum_{\ell'' m''} g_{\ell''} Y^*_{\ell'' m''}(\hat{\boldsymbol{n}}_0) \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \\ &\quad \times \begin{pmatrix} \ell & \ell' & \ell'' \\ -m & m' & m'' \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix} (-1)^m, \end{aligned} \tag{9}$$

where relation (B3) for Wigner 3j Symbols was used. Using this expression, the relation $\langle a^*_{\ell m} a_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}$ and the orthogonality of Wigner symbols (equation B1), one can write $\langle \tilde{C}_\ell \rangle$ as,

$$\langle \tilde{C}_\ell \rangle = \sum_{\ell'} C_{\ell'} K(\ell, \ell'). \tag{10}$$

With C_ℓ we will always mean $\langle C_\ell \rangle$ when we are referring to the full-sky power spectrum. In this expression, $K(\ell, \ell')$ is the Gabor kernel,

$$K(\ell, \ell') = (2\ell'+1) \sum_{\ell''} g^2_{\ell''} \frac{(2\ell''+1)}{(4\pi)^2} \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2. \tag{11}$$

The Legendre coefficients g_ℓ, are found by the inverse Legendre transformation,

$$g_\ell = 2\pi \int_{\theta=0}^{\theta=\theta_C} G(\theta) P_\ell(\cos\theta)\, \mathrm{d}\cos\theta, \tag{12}$$

where θ_C is the cut-off angle where the window goes to zero. One sees from the expression for the kernel that there is no dependency on $\hat{\boldsymbol{n}}_0$. This means that $\langle \tilde{C}_\ell \rangle$ is the same, independent of where the Gabor window is centred. In the rest of this section we will study the shape of this kernel which couples the $\tilde{C}_\ell$ on the apodized sphere.

In Fig. 1 we have plotted the kernel for a Gaussian Gabor window,

$$G(\theta) = \mathrm{e}^{-\theta^2/(2\sigma^2)} \quad \theta \leqslant \theta_C, \tag{13}$$

$$G(\theta) = 0 \qquad \theta > \theta_C, \tag{14}$$

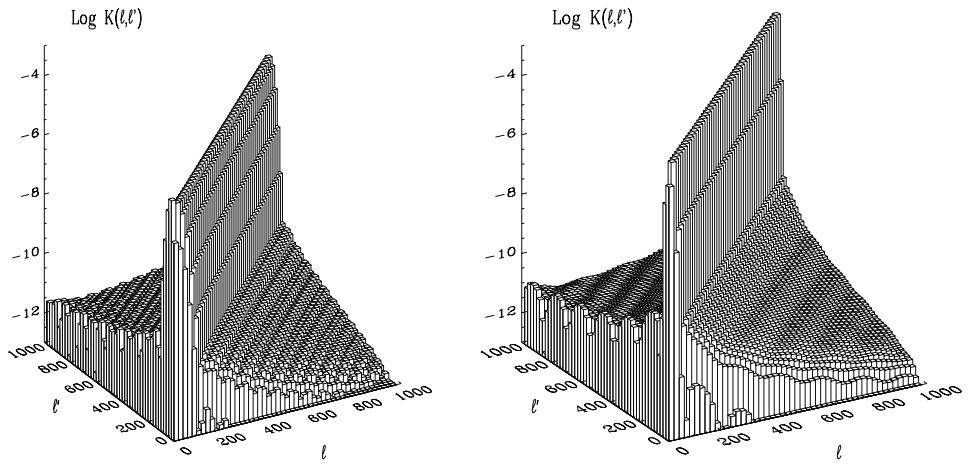

Figure 1. The logarithm of the kernel $K(\ell, \ell')$, describing the connection between the spherical harmonic coefficients C_ℓ on the full sky and the corresponding coefficients $\tilde{C}_\ell$ on the apodized sky via the relation $\tilde{C}_\ell = \sum_{\ell'} K(\ell, \ell') C_{\ell'}$. The figure shows the kernel for a 5° and a 15° FWHM Gaussian Gabor window with $\theta_C = 3\sigma$ (left and right, respectively).

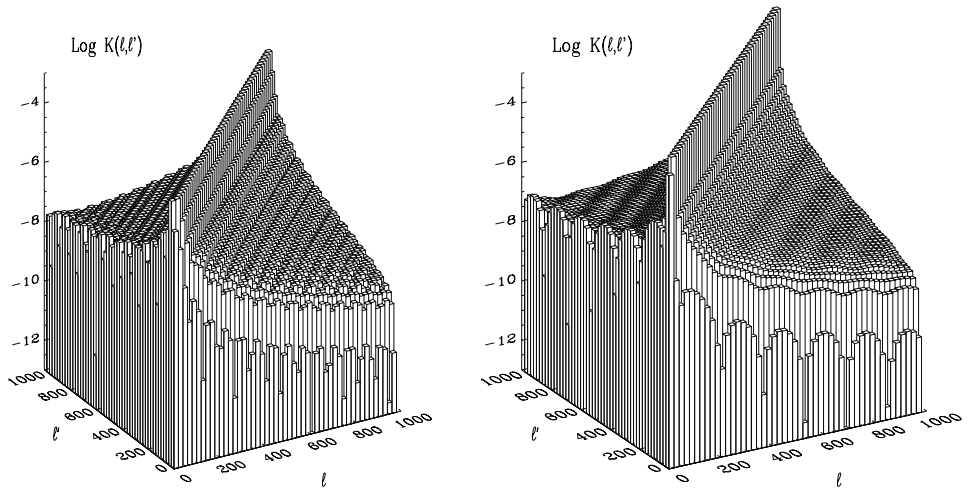

Figure 2. Same as Fig. 1 for top-hat Gabor windows covering the same area on the sky as the Gaussian windows.

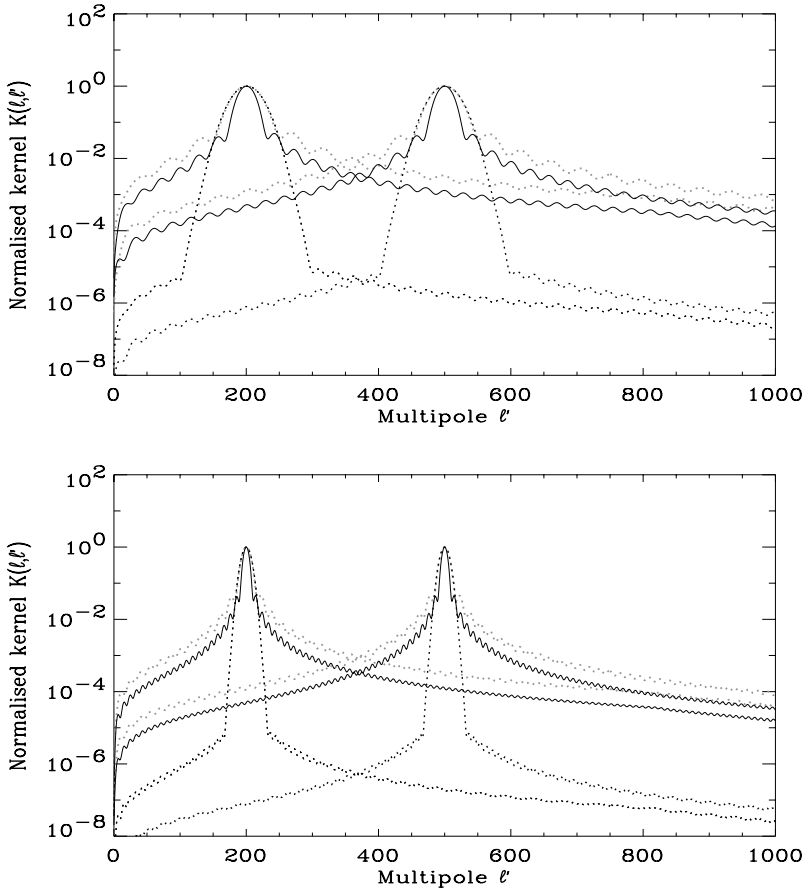

Figure 3. The panels show slices of the kernel $K(\ell, \ell')$ connecting the full sky and cut sky spherical harmonic coefficients. The full kernels are shown in Figs 1 and 2. The slices are taken at $\ell = 200$ and $\ell = 500$ for the 5° (upper plot) and the 15° (lower plot) degree Gaussian Gabor window (dotted black lines). The solid lines are for the corresponding (same area on the sky) top-hat window $W_{\rm A}$. The dotted grey lines are for the top-hat window $W_{\rm I}$ having the same integrated area as the Gaussian window. The kernels are here normalized so that the peak in the given slice has its maximum at 1. In this way one can easier compare the shape of the kernels.

with *5* and 15 degrees full width at half maximum (FWHM; corresponding to $\sigma = 2.12^\circ$ and $\sigma = 6.38^\circ$) and $\theta_{\rm C} = 3\sigma$. One sees that the kernel is centred about $\ell = \ell'$, and falls off rapidly. Fig. 2 shows the same for the corresponding top-hat Gabor windows,

$$G(\theta) = 1 \quad \theta \leqslant \theta_{\rm C}, , \tag{15}$$

$$G(\theta) = 0 \quad \theta > \theta_{\rm C}. \tag{16}$$

The top-hat windows are covering the same area on the sky as the corresponding Gaussian windows in Fig. 1 ($\theta_{\rm C}$ is the same). Ones sees that the diagonal is broader for the smaller windows indicating stronger couplings. Another thing to notice is that whereas the

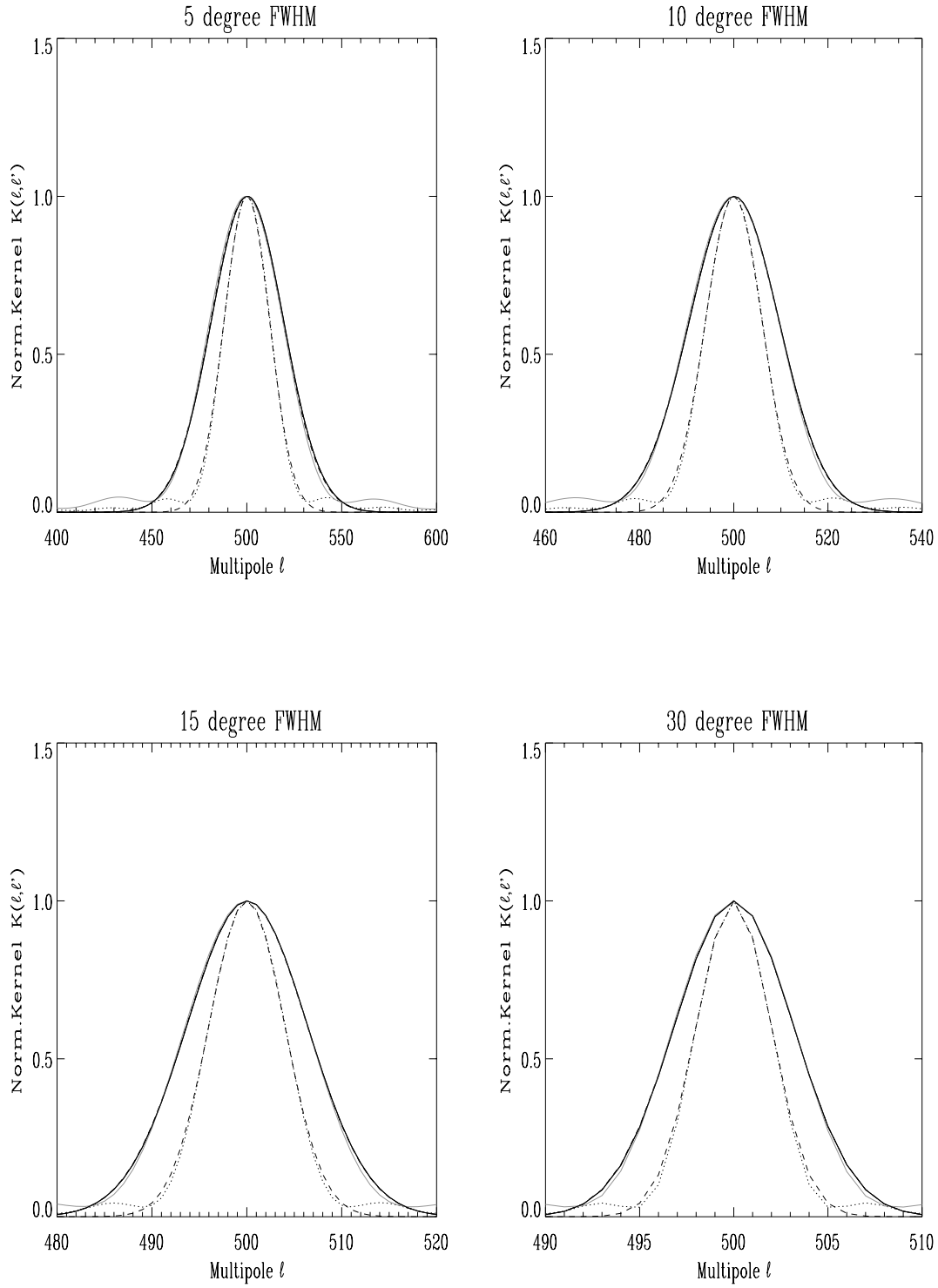

Figure 4. The panels show a slice of the kernel $K(\ell, \ell')$ connecting the full-sky and cut-sky spherical harmonic coefficients. The slices are taken at $\ell = 500$ for a 5°, 10°, 15° and 30° FWHM Gaussian Gabor window (solid line) with a $\theta_C = 3\sigma$ cut-off. The dotted line shows the kernel for a top-hat window W_A covering the same area on the sky. The grey lines, which are almost on top of the lines for the Gaussian Gabor windows, show the kernels for a top-hat window W_I, which has the same integrated area as the Gaussian windows W_G. The dashed lines, which are almost on top of the dotted and solid lines (and for this reason not so easily seen in the plot), are Gaussian fits to the curves.

kernel for the top-hat Gabor window only falls by about 4 orders of magnitude from the diagonal to the far off-diagonal elements, the Gaussian Gabor kernel falls by about 8 orders of magnitude (the vertical axis on the four plots are the same). The smooth cut-off of the Gaussian Gabor window cuts off long range correlations in spherical harmonic space. One of the aims of the first part of this paper is to see how the pseudo power spectrum of a given disc on the sky (top-hat window) is affected by the multiplication with a Gaussian Gabor window. For this reason the pseudo-spectrum will be studied for a top-hat and a Gaussian covering the same area on

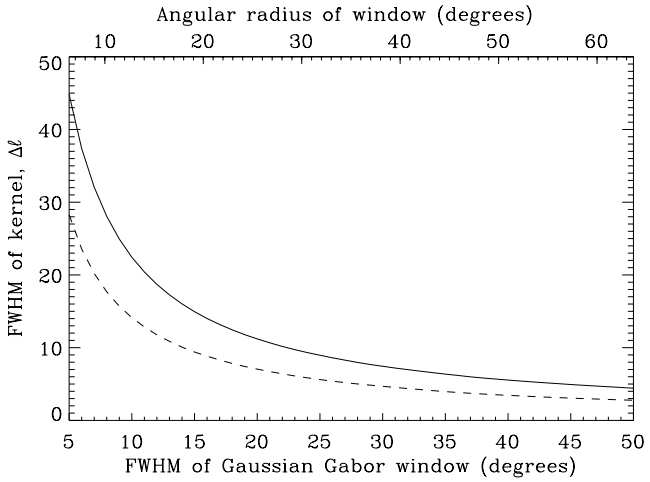

Figure 5. The figure shows the uncertainty relation $\Delta\ell\Delta\theta = \text{constant}$ for a Gabor transform on the sphere. The solid line shows the width $\Delta\ell$ of the Gabor kernel $K(\ell, \ell')$ connecting the full-sky and the cut-sky power spectra when applying a Gaussian Gabor window with a cut $\theta_C = 3\sigma$. The FWHM is shown on the lower abscissa. The dashed line shows the width of the kernel for a top-hat window. The full radius of the top-hat window is shown on the upper abscissa. The curves are well described by $\Delta\ell = 220/\theta_{\rm FWHM}$ and $\Delta\ell = 175/\theta_{\rm radius}$ for the Gaussian and top-hat windows respectively. As will be discussed in Section 3.2, this relation also describes the width of the correlation matrix of $\tilde{C}_\ell$. The width of the correlation matrix using a Gaussian window follows the relation in this plot times a factor 1.42. For the top-hat window, the width of the correlation matrix is the same as for the kernel shown in this figure.

the sky. We will also study a top-hat window which has the same integrated area as the Gaussian window. The cut-off angle $\theta_C = \theta_{\rm int}$ for these windows is given by

$$\int_{\theta=0}^{\theta=3\sigma} G(\theta)\,{\rm d}\cos\theta = \int_{\theta=0}^{\theta=\theta_{\rm int}} {\rm d}\cos\theta. \qquad (17)$$

In this section we will be comparing a Gaussian window (called W_G) having $\theta_C = 3\sigma$ with a top-hat window (called W_A) having the same area on the sky ($\theta_C = 3\sigma$) and with a top-hat window (called W_I) having the same integrated area ($\theta_C = \theta_{\rm int}$).

In Fig. 3, we have plotted slices of the kernel at $\ell = 200$ and $\ell = 500$ for the 5° and 15° FWHM Gaussian Gabor windows (dashed line). The solid line is the corresponding kernel (same area on the sky) when using the top-hat Gabor window (W_A). One sees that the Gaussian window effectively cuts off long range correlations whereas the top-hat window is narrower close to the diagonal. The Gaussian window has larger short range correlations. The coloured lines show the slice of the kernel for a top-hat window having the same integrated area as the Gaussian window (W_I). These kernels have the same widths as the kernels for the Gaussian windows, but the long-range correlations are significantly larger. In (Hobson & Magueijo 1996) it was shown that in the flat-sky approximation, the long-range correlations are significant when the window has a sharp cut-off. On the sphere we see that even for a sharp top-hat window the long-range correlations are damped.

Fig. 4 shows how the width of the kernel gets narrower and the correlations smaller as the Gabor window opens up. The four kernels are shown for $\ell = 500$ and the Gaussian windows have 5°, 10°, 15° and 30° FWHM with $\theta_C = 3\sigma$. The same kernels for the top-hat windows W_A (dotted lines) and W_I (coloured line) are plotted on top. Gaussian fits are plotted on top of the kernels and show that the kernels are very close to Gaussian functions near the diagonal.

In Fig. 5 we have plotted the relation between the FWHM width $\Delta\ell$ of the kernel and the size $\Delta\theta$ of the window for Gaussian and top-hat windows. The two curves are very well described by $\Delta\ell = 220/\theta_{\rm FWHM}$ for the Gaussian window ($\theta_{\rm FWHM}$ in degrees) and $\Delta\ell = 175/\theta_{\rm radius}$ ($\theta_{\rm radius}$ being the radius of the top-hat window in degrees) for the top-hat window. Clearly for a given observed area of the sky, multiplying with a Gaussian will increase the FWHM of the kernel. This is also what was seen in Figs 3 and 4. We will see that this results in a lower spectral resolution for the Gaussian window compared to the top-hat window. However, the lower long-range correlations of the Gaussian window makes the shape of the pseudo power spectrum closer to that of the full-sky power spectrum.

In Fig. 6, we show the shapes of the $\tilde{C}_\ell$ for Gaussian and top-hat windows compared to the full-sky spectrum. The plots which were made using the analytical formula (10) show $\tilde{C}_\ell$ for a 5° and 15° FWHM Gaussian Gabor window (solid line) cut at 3σ. The corresponding spectrum for the top-hat Gabor window W_A is shown as dotted lines and for the top-hat window W_I as coloured lines. The spectra are normalized in such a way that they can be compared to the full-sky power spectrum (dashed line). For the 5° FWHM window one can still distinguish the four lines. At this window size the pseudo-spectra are very similar to the full-sky spectra, but with small deviations depending on the shapes of the kernel and the power spectrum. In this case the spectrum for the Gaussian window seems to be smaller at the peaks and larger at the troughs whereas the spectrum for the top-hat windows is always larger.

For the 15° FWHM windows the pseudo-spectrum using the Gaussian Gabor window are on top of the full-sky power spectrum. For the top-hat windows it is still possible to distinguish the pseudo-spectrum from the full-sky power spectrum although the lines are still very close. The plot implies that the $\tilde{C}_\ell$ could be good estimators of the underlying full-sky C_ℓ provided that the window is big enough. Note that for small windows, the Gaussian Gabor window makes the pseudo-spectrum a better estimator than the pseudo-spectrum for a top-hat window at higher multipoles. In (Hobson & Magueijo 1996) it was shown in the flat-sky approximation that the pseudo power spectrum for small fields get significantly distorted, but that the shape of the pseudo-spectrum gradually approaches the shape of the full-sky power spectrum when the window gets larger. We see here that the same results applies to the treatment on the sphere. In the flat-sky approximation however, the error in estimating the average power spectrum from the pseudo power spectrum from one single realization is bigger due to the long range correlations of the pseudo power spectrum coefficients in the flat-sky approximation.

One feature which is very prominent is the additional peak at low ℓ for the Gaussian window. The reason for this peak comes from the fact that the diagonal in the Gaussian kernel is broader than in the top-hat kernel for a top-hat window with corresponding area. For the low multipoles the power spectrum is dropping rapidly because of the Sachs–Wolfe effect and the lowest multipole C_ℓ are much bigger than the C_ℓ for higher multipoles. Because the Gaussian kernel is broad, the $\tilde{C}_\ell$ at low multipoles will pick up more from the C_ℓ at lower multipoles than the narrower top-hat kernel (see Fig. 3). These low multipole C_ℓ have very high values compared to the higher multipole C_ℓ and for that reason the $\tilde{C}_\ell$ for the Gaussian window will get a higher value. This is illustrated in Fig. 7 where a slice of the kernel at $\ell = 50$ is shown for the 5° FWHM Gaussian Gabor window (solid line) and the corresponding top-hat W_A (dashed line) normalized to one at the peak. The dotted line shows a typical power

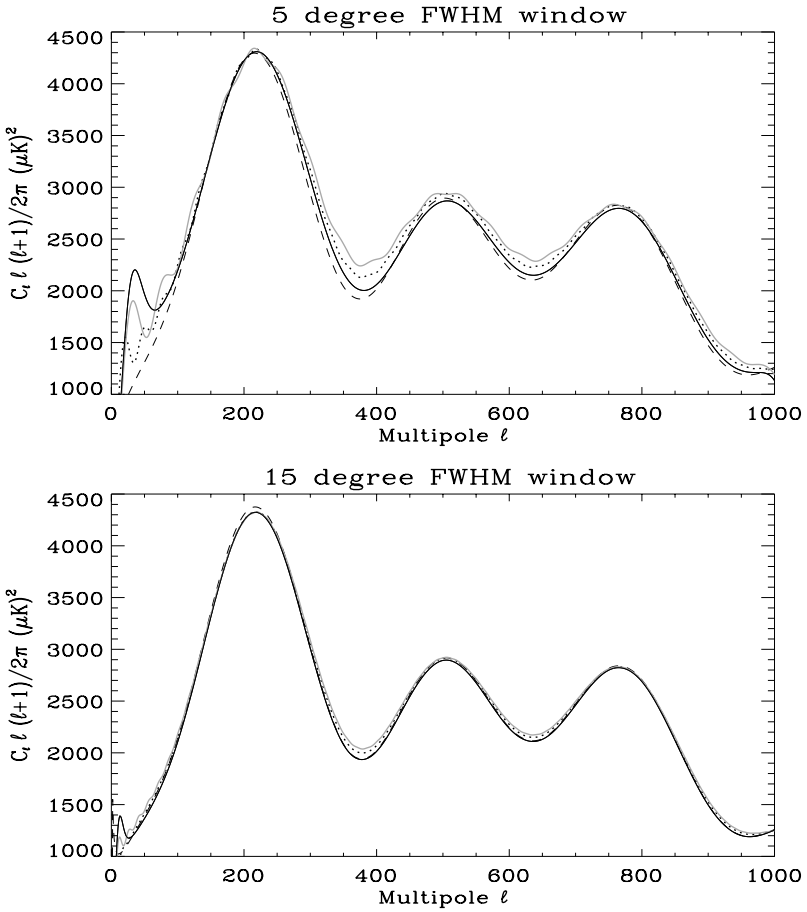

Figure 6. The windowed power spectra $\tilde{C}_\ell$ for a *5* and 15° FWHM Gaussian Gabor window $W_{\rm G}$ cut at $\theta_{\rm C} = 3\sigma$ (solid line) and for a top-hat window $W_{\rm A}$ covering the same area on the sky (dotted line). The spectrum for the top-hat window $W_{\rm I}$ for which the integrated area of the window corresponds to the Gaussian is shown as a grey line. All spectra are normalized in such a way that they can be compared directly with the full-sky spectrum, which is shown on each plot as a dashed line. Only in the first plot are all four lines visible. In the last plot, the full-sky spectrum and the Gaussian pseudo-spectrum (dashed and solid lines) are only distinguishable in the first few multipoles.

spectrum. Clearly the Gaussian kernel will pick up more of the high-value C_ℓ at low multipoles. Note that for the pseudo-spectrum for the top-hat window $W_{\rm I}$, where the integral of the top-hat window corresponds to the integrated Gaussian window (grey line), there is also a peak at low multipole. The reason is that the width of the kernel is the same as for the Gaussian.

In Fig. 8 we show the pseudo power spectra for a particular realization using a 15° FWHM Gaussian window (upper plot) and a top-hat window $W_{\rm A}$(lower plot). The pseudo-spectra are compared to the average full-sky spectra shown as a dashed line. The dark shaded area shows the expected 1σ cosmic and sample variance on the pseudo-spectra taken from the formulae to be developed in the next sections. The lighter shaded area shows only cosmic variance. Note that the pseudo-spectrum for the Gaussian window is smoother than the pseudo-spectrum for the top-hat window. This shows the lower spectral resolution of the Gaussian window due to the broader kernel.

3 LIKELIHOOD ANALYSIS

In this section we will show how the pseudo power spectrum can be used as input to a likelihood analysis for estimating the full-sky power spectrum from an observed disc on the sky multiplied with a Gabor window. We will in this section concentrate on a Gaussian Gabor $W_{\rm G}$ window, but the formalism is valid for any azimuthally symmetric Gabor window. We start by showing that the pseudo-C_ℓ are close to Gaussian distributed which allows for a Gaussian form of the likelihood function. Then we show how the correlation matrices can be calculated quickly for an axisymmetric patch on the sky with uncorrelated noise. The extension to the more realistic situation with correlated noise and non-axisymmetric sky patches will be made in the next section. We will show the results of power spectrum estimations with different noise profiles and window sizes. We will also show that the use of a window different from the top-hat window can be advantageous for some noise profiles, even if the window has a lower spectral resolution than the top-hat window.

3.1 The form of the likelihood function

To know the form of the likelihood function, one needs to know the probability distribution of $\tilde{C}_\ell$. In Figs 9 and 10 we show the probability distribution from 10 000 simulations with a 5° and 15° FWHM Gaussian Gabor window respectively. The dashed line shows a Gaussian with mean value and standard deviation found from the formulae given in the previous and next section. One can see that the probability distribution is slightly skewed for low ℓ, but for high ℓ it seems to be very well approximated by a Gaussian. Also, the small window shows more deviations from a Gaussian than the

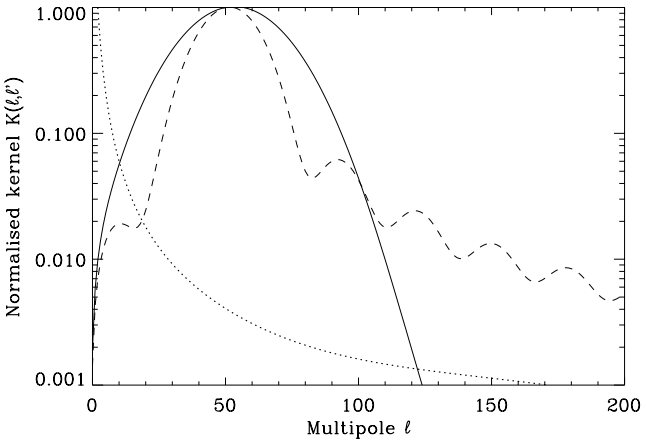

Figure 7. The figure shows a slice of the kernel $K(\ell, \ell')$ connecting the full-sky and cut-sky spherical harmonic coefficients. The slice is taken at $\ell = 50$ for a 5° FWHM Gaussian Gabor window (solid line) and a corresponding top-hat window W_A (dashed line). The kernels are normalized to one at the peak. A typical power spectrum normalized to one at the quadrupole is plotted as a dotted line. The figure aims at explaining the extra peak in the pseudo power spectrum at low multipoles for the Gaussian Gabor window shown in Fig. 6.

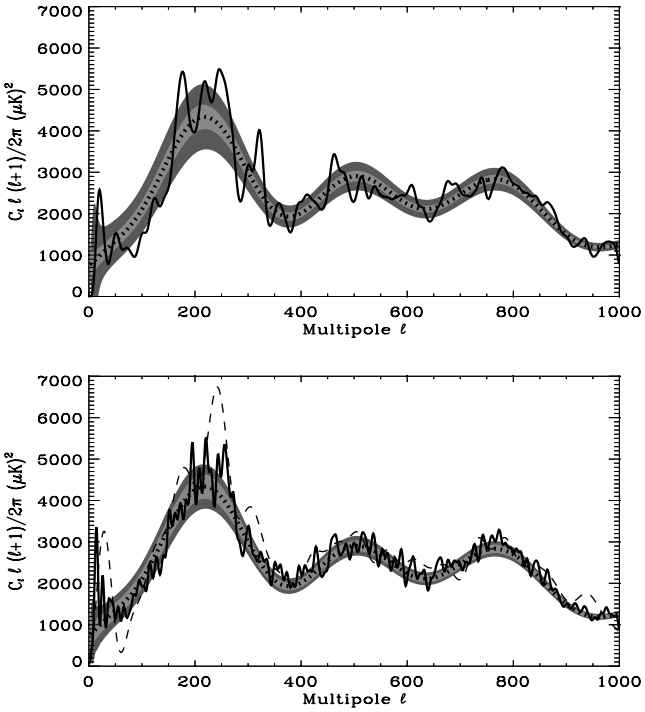

Figure 8. One realization of the windowed power spectra. The upper plot shows a realization of a pseudo power spectrum using a 15 degree FWHM Gaussian Gabor window. The pseudo-spectrum is normalized in such a way that it can be compared directly to the full-sky spectrum, the average of which is is shown as a dotted line. The lower plot shows the same realization using a corresponding top-hat window. The light shaded area shows 1σ cosmic variance around the full sky average spectrum. The darker area shows 1σ cosmic and sampling variance taken from the theoretical formula. On the lower plot, the pseudo-spectrum with a top-hat window W_A with the same integrated area as in the upper plot is shown as a dashed line.

bigger window. In Fig. 11 we show that this result is not limited to the Gaussian window. The plot shows the probability distribution from a simulation with a top-hat Gabor window W_A covering the same area on the sky as the 15° FWHM Gaussian window. For this window the probability distribution is also close to Gaussian.

From the above plots it seems to be reasonable to approximate the likelihood function with a Gaussian, provided the window is big enough and multipoles at high enough ℓ values are used,

$$\mathcal{L} = \frac{\mathrm{e}^{-1/2\boldsymbol{d}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{d}}}{\sqrt{2\pi \det \mathbf{M}}}. \tag{18}$$

Omitting all constant terms and factors, the log-likelihood can then be written:

$$L = \boldsymbol{d} \tag{19}$$

Here $\boldsymbol{d}$ is the data vector which contains the observed $\tilde{C}_\ell$ for the set of sample ℓ-values ℓ_i. The data is taken from the observed windowed sky in the following way:

$$d_i = \tilde{C}_{\ell_i} - \left\langle \tilde{C}_{\ell_i} \right\rangle. \tag{20}$$

The matrix $\mathbf{M}$ is the covariance between pseudo-C_l, the elements of which are given by:

$$M_{ij} = \left\langle \tilde{C}_{\ell_i} \tilde{C}_{\ell_j} \right\rangle - \left\langle \tilde{C}_{\ell_i} \right\rangle \left\langle \tilde{C}_{\ell_j} \right\rangle. \tag{21}$$

In Appendices (E) and (F), later, we have found expressions which enable fast evaluations of d_i and M_{ij} for signal and noise. These major results are are given in equations (E8), (F13) and (F21) and the recursion which enables fast calculation of these expressions is given in equation (C15). In the derivations of the expressions for M_{ij}, the rotational invariance of the (non-averaged) $\tilde{C}_\ell$ shown in Appendix (D) was used. Because of this rotational invariance, all derivations can be done with the Gabor window centred at the north pole. In Fig. 12 we used the full formula (equation E8) from Appendix (E) to calculate the signal correlation matrix for a typical power spectrum with a 15° FWHM Gaussian Gabor window (note that in the figure, the correlation matrix is normalized with the pseudo power spectrum). The correlation between $\tilde{C}_\ell$ of different multipoles is falling of rapidly with the distance from the diagonal. Only exception being the small 'wall' at low multipoles which again comes from the coupling to the smallest multipoles which have very high values.

3.2 Likelihood estimation and results

Because of the limited information content in one patch of the sky one can not estimate the full sky C_ℓ for all multipoles ℓ. For this reason the full-sky power spectrum has to be estimated in N^{bin} bins. Also the algorithm to minimize the log-likelihood needs the different numbers to be estimated to be of roughly the same order of magnitude. For this reason we estimate for some parameters D_b, which for bin b is defined as

$$C_\ell = \frac{D_b}{\ell(\ell+1)}, \ell_b \leqslant \ell < \ell_{b+1}, \tag{22}$$

where ℓ_b is the first multipole in bin b.

As the $\tilde{C}_\ell$ are coupled, one cannot use all multipoles in the data vector, and so the covariance matrix would in this case become singular. One has to choose a number N^{in} of multipoles ℓ_i for which one finds d_i. How many multipoles to use depends on how tight the $\tilde{C}_\ell$ are coupled which depends on the width $\Delta\ell_{\mathrm{kern}}$ of the kernel (Fig. 5) or the width $\Delta\ell_{\mathrm{cor}}$ of the correlation matrix. The width of the correlation matrix(normalized with the pseudo power spectrum) varies with window size in the same way as the width of the kernel

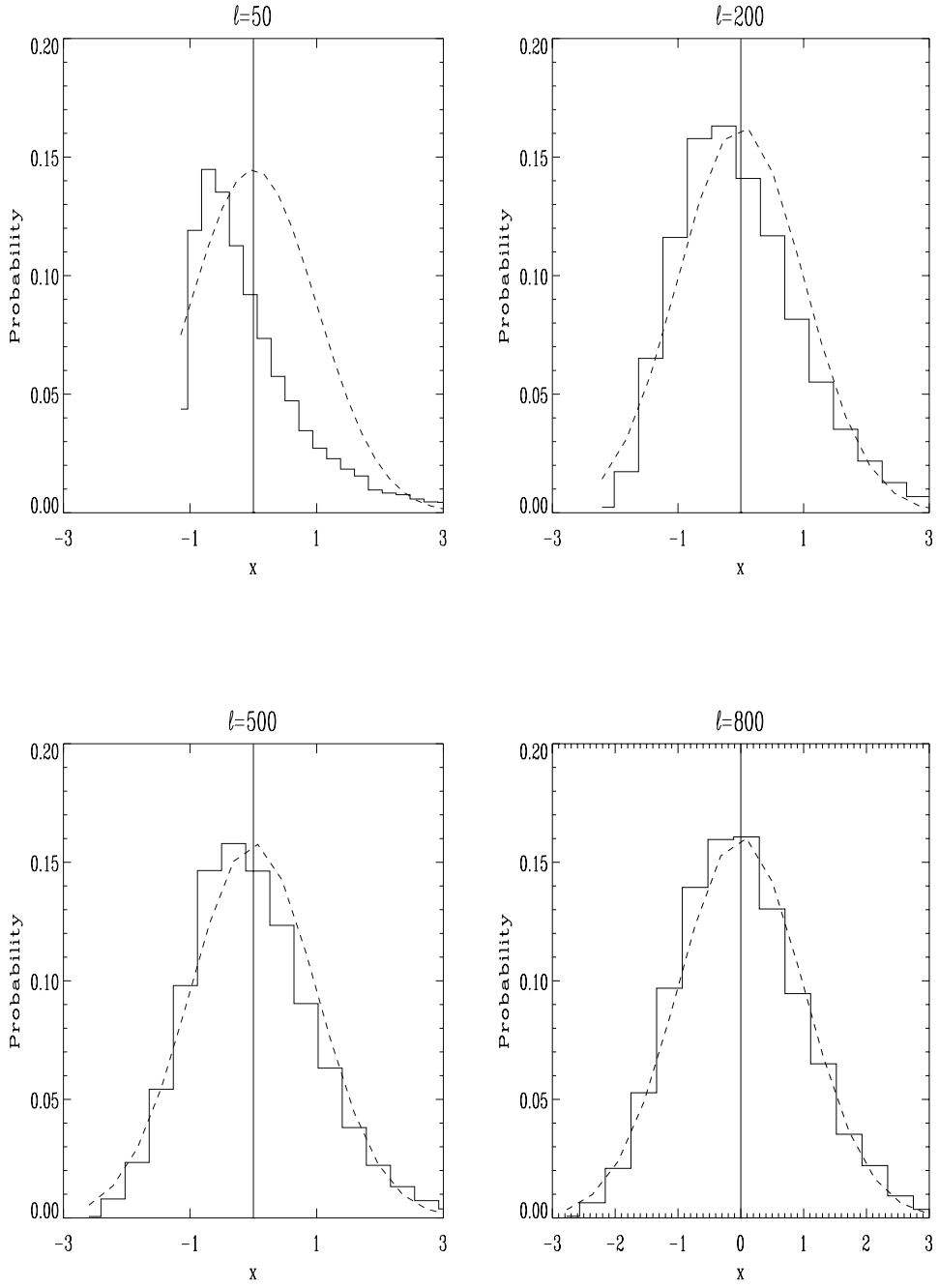

Figure 9. The probability distribution of $\tilde{C}_\ell$ taken from 10 000 simulations with a 5° FWHM Gaussian Gabor window truncated at $\theta_C = 3\sigma$. The variable x is given as $x = (\tilde{C}_\ell - \langle\tilde{C}_\ell\rangle)/\sqrt{\langle(\tilde{C}_\ell - \langle\tilde{C}_\ell\rangle)}$. The dashed line is a Gaussian with the theoretical mean and standard deviation of the $\tilde{C}_\ell$. The plot shows the $\tilde{C}_\ell$ distribution for $\ell = 50$, $\ell = 200$, $\ell = 500$, and $\ell = 800$. The probabilities are normalized such that the integral over x is 1.

varies with window size. In fact for the top-hat window these two widths are the same and for the Gaussian window we found that $\Delta\ell_{\rm cor} \approx 1.42\Delta\ell_{\rm kern}$. The optimal number of $N^{\rm in}$ to use seems to be $N^{\rm in} \approx 3/2\ell_{\rm max}/\Delta\ell_{\rm kern}$. To use a lower $N^{\rm in}$ increases the error bars on the estimates and a higher $N^{\rm in}$ does not improve the estimates. One can at most fit for as many C_ℓs as the number of $\tilde{C}_\ell$ ($N^{\rm in}$) one has used in the analysis. Therefore one needs to find a number $N^{\rm bin} \leqslant N^{\rm in}$ of bin values D_b from which one can construct the full-sky power spectrum C_ℓ.

In Appendices (E) and (F), later, we found that the full correlation matrix can be written as

$$M_{ij} = M^{\rm S}_{ij} + M^{\rm N}_{ij} + M^{\rm X}_{ij}, \tag{23}$$

where $M^{\rm N}_{ij}$ is the noise correlation matrix which has to be precomputed for a specific noise model (analytically or by Monte Carlo as will be shown in Section 4). The signal and signal–noise cross-correlation matrices are of the form

$$M^{\rm S}_{ij} = \sum_b \sum_{b'} D_b D_{b'} \chi(b, b', i, j), \tag{24}$$

$$M^{\rm X}_{ij} = \sum_k D_b \chi'(b, i, j), \tag{25}$$

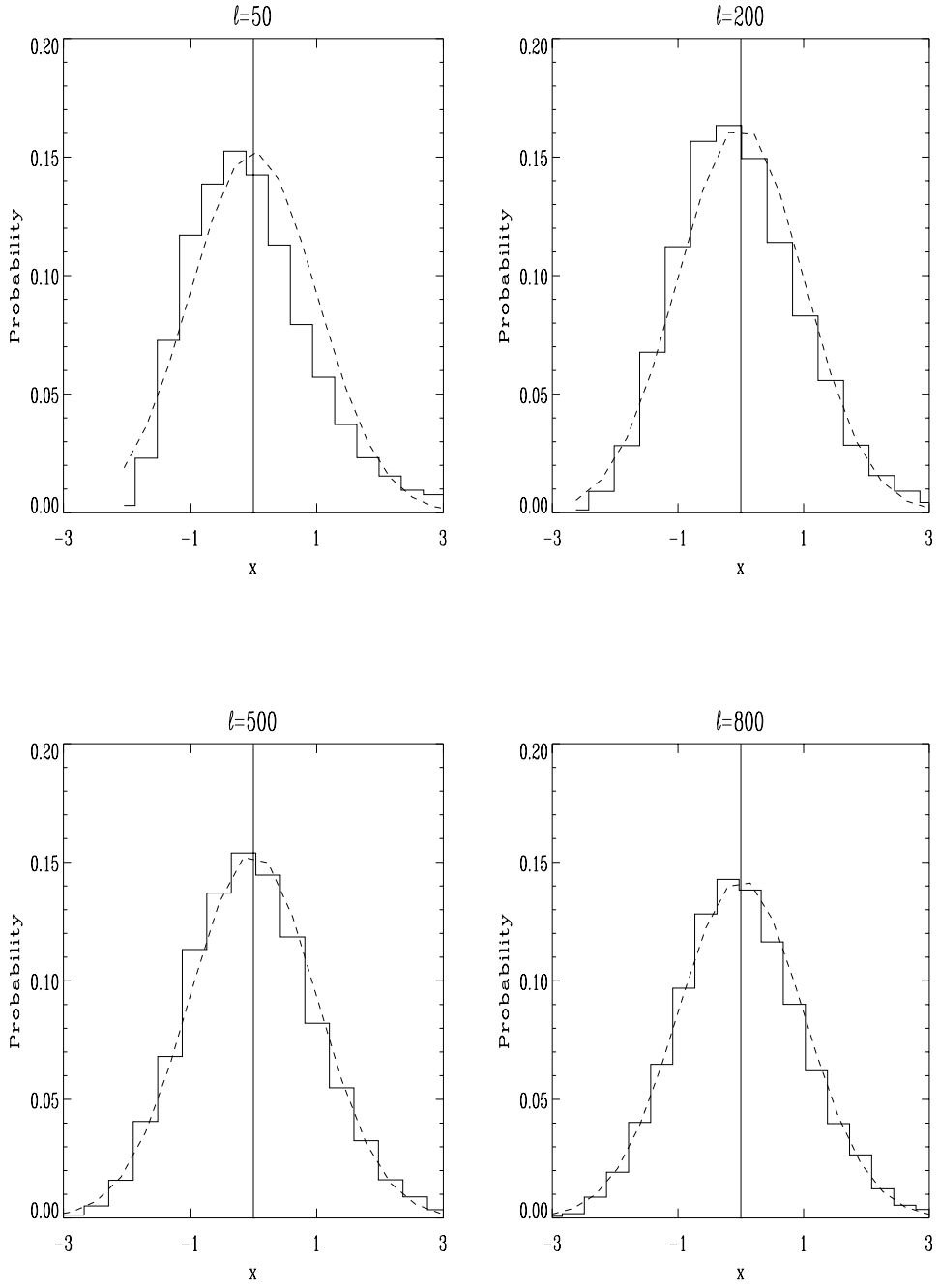

Figure 10. Same as Fig. 9 but for a 15° FWHM Gaussian Gabor window.

where the χ-functions can be precomputed using formulae (E10) and (F26).

We will now describe some test simulations to show how the method works. As a first test, we used the same model as was used in (Hivon et al. 2002), with $\Omega_{\rm total} = 1$, $\Omega_\Lambda = 0.7$, $\Omega_{\rm b}h^2 = 0.03$ and $n_{\rm s} = 0.975$. These are the parameters from the combined Maxima–Boomerang analysis (Jaffe et al. 2001). We used a circular patch with 15.°5 radius covering the same fraction of the sky as in (Hivon et al. 2002). Using HEALPIX we simulated a CMB sky using a standard cold dark matter (CDM) power spectrum with $l_{\rm max} = 1024$ and a 7-arcmin pixel size ($N_{\rm side} = 512$ in HEALPIX language). We smoothed the map with a 10-arcmin beam and added non-correlated non-uniform noise to it. Here a Gaussian Gabor window with FWHM $= 12^\circ$ was used with a cut-off $\theta_{\rm C} = 3\sigma$. For the likelihood estimation, we had $N^{\rm bin} = 20$ full sky C_ℓ bins and $N^{\rm in} = 100$ $\tilde{C}_\ell$ values between $\ell = 2$ and $\ell = 960$. In Fig. 13 one can see the result. The shaded areas are the expected 1σ variance with and without noise. These were found from the theoretical formula

$$\Delta C_b = \sqrt{\frac{2}{\nu_b}}(C_b + N_b), \tag{26}$$

where N_b is the noise 'on the sky', ν_b is the effective number of degrees of freedom given as

$$(2\ell_b + 1)\Delta\ell f_{\rm sky}\frac{w_2^2}{w_4}, \tag{27}$$

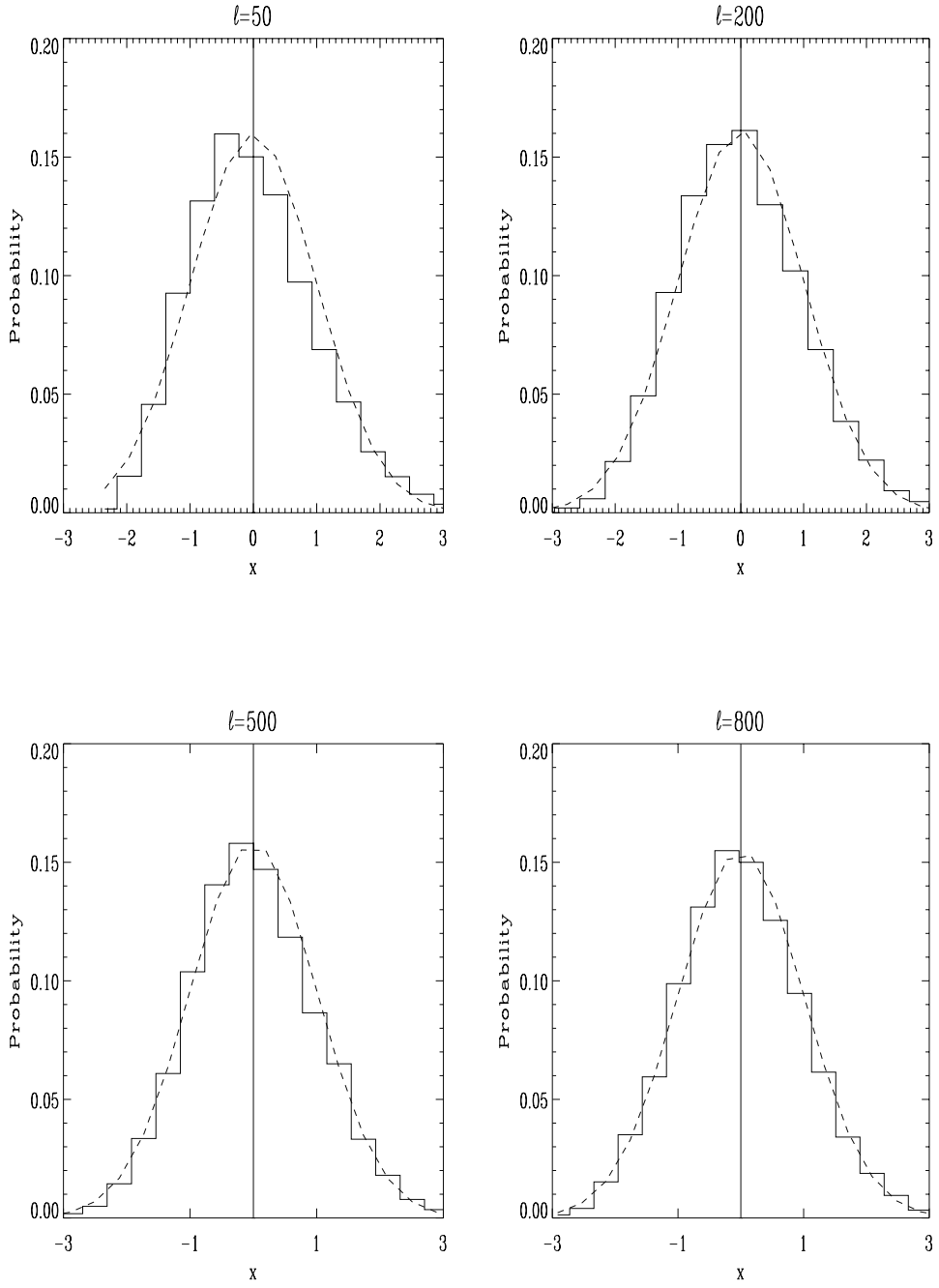

Figure 11. Same as Fig. 10 but for a top-hat Gabor window $W_{\rm A}$ covering the same area on the sky.

and the w_i factors are dependent on the window according to

$$f_{\rm sky} w_i = \frac{1}{4\pi} \int_{4\pi} {\rm d}\hat{\boldsymbol{n}} G^i(\hat{\boldsymbol{n}}). \tag{28}$$

This formula is exact for a uniform noise model (Hivon et al. 2002) and is similar to the one used in most publications. It is in this case a very good approximation even with non-uniform noise. In the next example, however, we will show that the formula has to be used with care. In the figure, the error bars on the estimates are taken from the Fisher matrix and the signal-to-noise ratio (S/N) = 1 at $\ell = 575$.

In Fig. 14, we have plotted the average of 1000 such simulations, with different noise and sky realizations. From the plot, the method seems to give an unbiased estimate of the power spectrum bins D_b. For the lowest multipoles the estimates are slightly lower than the binned input spectrum. This is a result of the slightly skewed probability distribution of $\tilde{C}_\ell$ for small windows at these low multipoles (see Figs 9 and 10). The probability that the $\tilde{C}_\ell$ at lower multipoles have a value lower than the average $\langle \tilde{C}_\ell \rangle$ is high, and the assumption about a Gaussian distribution about this average leads the estimates to be lower. When a bigger area of the sky is available such that several patches can be analysed jointly to give the full-sky power spectrum, this bias seems to disappear. This will be shown in Section (4.1).

In this example one can see that the 1σ error bars from the Monte Carlo analysis coincide very well with the theoretical error shown

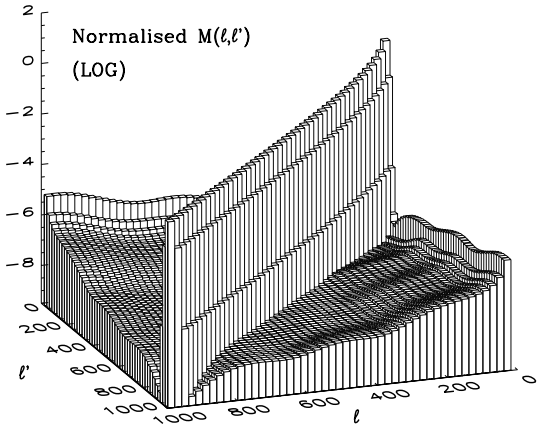

Figure 12. The figure shows the correlation matrix $M_{\ell\ell'}$ between pseudo-spectrum coefficients normalized with the pseudo-spectrum $(\langle\tilde{C}_\ell^{\rm T}\tilde{C}_{\ell'}^{\rm T}\rangle - \langle\tilde{C}_\ell^{\rm T}\rangle\langle\tilde{C}_{\ell'}^{\rm T}\rangle)/(\langle\tilde{C}_\ell^{\rm T}\rangle\langle\tilde{C}_{\ell'}^{\rm T}\rangle)$ for a 15° FWHM Gaussian Gabor window. A standard CDM power spectrum was used to produce this matrix.

as shaded areas from the formula in Hivon et al. (2002). Note that the error bars on the higher ℓ are smaller than in Hivon et al. (2002) because the noise model used in that paper was not white. Also, they took into account errors due to map making which is not considered here.

As a next test, we used a simulation with the same resolution and beam size. The power spectrum was this time a standard CDM power spectrum. We used an axisymmetric noise model with noise increasing from the centre and outwards to the edges (see Fig. 15). This is the kind of noise model which could be expected from an experiment scanning on rings, with the rings crossing in the centre. We now use a circular patch with 18.5 radius and a FWHM = 15° Gaussian Gabor window cut at $\theta_{\rm C} = 3\sigma$. An interesting point now is that the Gabor window is decreasing from the centre and outwards, which is the opposite of the noise pattern. This gives the pixels with low noise high significance in the analysis and the pixels with high noise low significance. One sees from the expressions for the signal and noise pseudo power spectra that the Gabor window will work differently on both. This means that S/N is different depending on the Gabor window. For this case, we have plotted the average pseudo power spectrum for signal and noise separately in Fig. 16. This shows the described effect. The S/N ratio is much higher for the Gaussian Gabor window in this case, favouring the use of this window for the analysis.

For this example we used $N^{\rm bin} = 20$ and $N^{\rm in} = 100$ again. The result is shown in Fig. 17. In Fig. 18 the average over 5000 simulations and estimations is shown. One can see that the estimate also does well beyond $\ell = 520$ which is where the effective S/N = 1. The method is still unbiased. The error bars in the part where noise dominates are here lower than the theoretical approximation (28) shown as the dark shaded area. The dashed lines show the theoretical 1σ variance taken from the inverse Fisher matrix which here gives a very good agreement with Monte Carlo.

In Fig. 19 we show the average (over 5000 estimations) of the correlation between the estimates D_b between different bins. The figure shows that the correlations between estimates are low and in fact in each line all off-diagonal elements are more than an order of magnitude lower than the diagonal element of that line. In Fig. 20 we show that the probability distribution of the estimates in Fig. 18 is almost Gaussian.

To test the method at higher multipoles we also did one estimation up to multipole $\ell = 2048$. We used HEALPIX resolution $N_{\rm side} = 1024$ and simulated a sky with a 8-arcmin Gaussian beam and added noise from a strongly varying non-uniform noise model. Both the beam and noise level were adjusted according to the specifications for the *Planck* High-Frequency Imager (HFI) 143 GHz

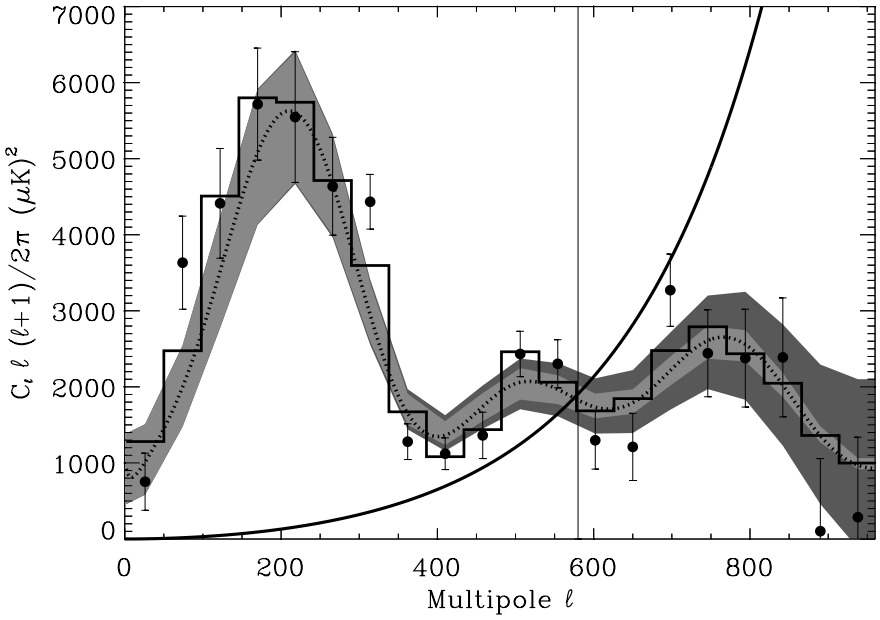

Figure 13. The analysis of an input model with $\Omega_{\rm total} = 1$, $\Omega_\Lambda = 0.7$, $\Omega_{\rm b}h^2 = 0.03$ and $n_{\rm s} = 0.975$. We used a non-uniform white noise model with S/N = 1 at $\ell = 575$. The dotted line is the input average full-sky power spectrum and the histogram shows the binned pseudo power spectrum for this realization (without noise). We used $N^{\rm bin} = 20$ bins and $N^{\rm in} = 100$ input sample points to the likelihood. The shaded areas around the binned average full-sky spectrum (which is not plotted) are the theoretical variance with and without noise. The light shaded area shows cosmic and sample variance and the darker shaded area also has variance due to noise inclusion. The variance due to noise was calculated using formula (26) for uniform noise. The 1σ error bars on the estimates are taken from the inverse Fisher matrix. The solid line increasing from the left to the right is the noise power spectrum.

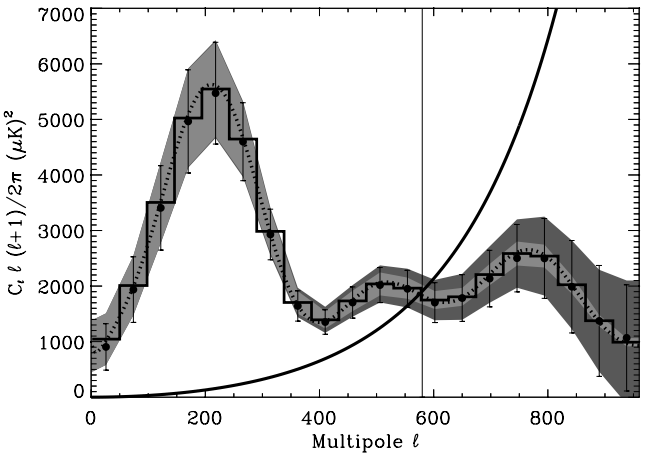

Figure 14. Same as in Fig. 13 but this is an average over 1000 simulations and estimations. The histogram is now the binned average full-sky power spectrum. The error bars on this plot are the 1σ variances taken from Monte Carlo simulations.

detector (Bersanelli et al. 1996). We used again a 15° FWHM Gaussian Gabor window cut 3σ away from the centre. In the estimation we used $N^{\rm bin} = 40$ bins and $N^{\rm in} = 200$ input $\tilde{C}_\ell$ between $\ell = 7$ and $\ell = 2048$. The average of 100 such simulations is shown in Fig. 21. Each complete likelihood estimation (which includes a total of about 25 likelihood evaluations) took about 8 min on a single processor on a 500 MHz DEC Alpha work station.

In Fig. 22, we have plotted the average of 300 estimations where the input data was the $\tilde{C}_\ell$ from simulations with a fixed CMB realization and varying noise realization. The dotted line shows the $N^{\rm in}$ $\tilde{C}_\ell$s (without noise) used as input to the likelihood. The histogram is as before the input pseudo-spectrum without noise binned in $N^{\rm bin}$ bins. This means that each histogram line shows the average of the dotted line over the bin. One can see that the estimated power spectrum is partly following the $N^{\rm in}$ input $\tilde{C}_\ell$ and partly following the binned power spectrum.

Finally, we made a comparison between a top-hat window and a Gaussian Gabor window. In this case we used uniform noise, so that the Gaussian and top-hat Gabor windows have the same S/N ratio which we set to *1* at $\ell = 520$. We used a disc with 18° radius, $N^{\rm in} = 200$ and $N^{\rm bin} = 20$. In Fig. 23 one can see the result. The lower plot shows the estimates with the Gaussian Gabor window (15° FWHM) and the upper with the top-hat window. The Gaussian window is suppressing parts of the data and for this reason gets a higher sample variance than the top-hat. This effect is seen in the plot. Clearly when no noise weighting is required the top-hat window seems to be the preferred window (which was also discussed in Hivon et al. 2002). This chapter has been concentrating on the Gaussian window to study power spectrum estimation in the presence of a window different from a top-hat. It has been shown that a different window can be advantageous when the noise is not uniformly distributed as one can then give data with different quality different significance.

4 EXTENSIONS OF THE METHOD

A real CMB experiment usually does not observe an axisymmetric patch of the sky. The noise between pixels is also usually correlated. In order to take these two issues into account we will discuss two extensions of the method. The formalism for the extensions

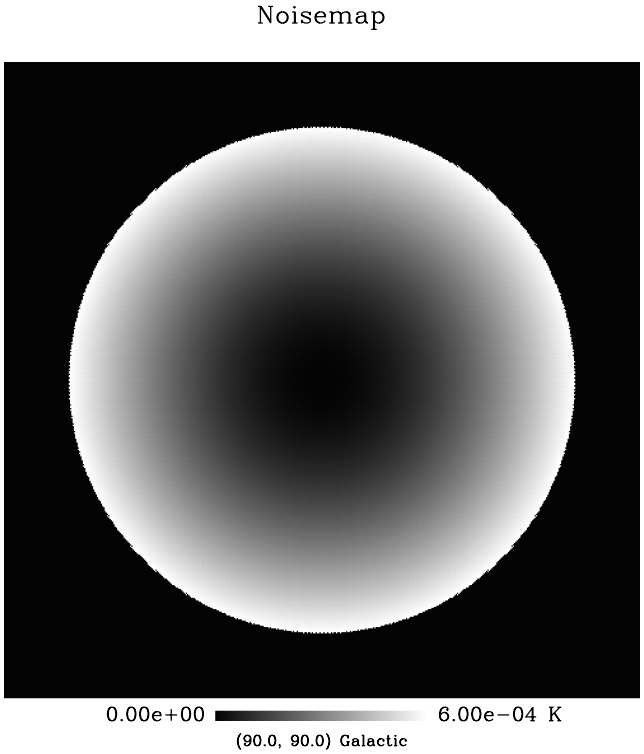

Figure 15. The noise map with noise increasing from the north pole and downwards. The figure shows a gnomic projection with the north pole in the centre.

are worked out and some simple examples are shown. Further investigations of these extensions are left for a future paper, where the analysis of *MAP* and *Planck* data will be discussed. To be able to analyse non-axisymmetric parts of the sky, we propose to split the area up into several axisymmetric pieces and use the pseudo-C_ℓ from all these patches in the data vector of the likelihood, and in this way analyse all patches jointly. We show that if the patches are not overlapping, the correlation between pseudo-C_ℓ from different patches is so weak that it can be neglected. To deal with correlated noise, we propose to use Monte Carlo simulations to find the noise correlation matrix. We demonstrate that for uncorrelated noise, one needs a few thousand simulations in order for the error bars on the C_ℓ estimates not to get larger than when using the analytic expression.

4.1 Multiple patches

It has been shown how one can carry out power spectrum estimation on one axisymmetric patch on the sky. The next question that arises is what to do when the observed area on the sky is not axisymmetric. In this case one can split the area into several axisymmetric pieces and use the $\tilde{C}_\ell$ from each piece. Then the $\tilde{C}_\ell$ from all the patches are used together in the likelihood maximization. The first thing to check before embarking on this idea is the correlation between $\tilde{C}_\ell$s in different patches. In Appendix (H) the analytical formula for the correlation matrix describing the correlations between $\tilde{C}_\ell$ for different patches was derived (equation H6 and H10). With these expressions we can check how the correlations decrease as the distance between the two patches increase.

After the expression (H6) was tested with Monte Carlo simulations, we computed the correlations between $\tilde{C}_\ell$ for two patches A and B where we varied the distance θ between the centres of A and

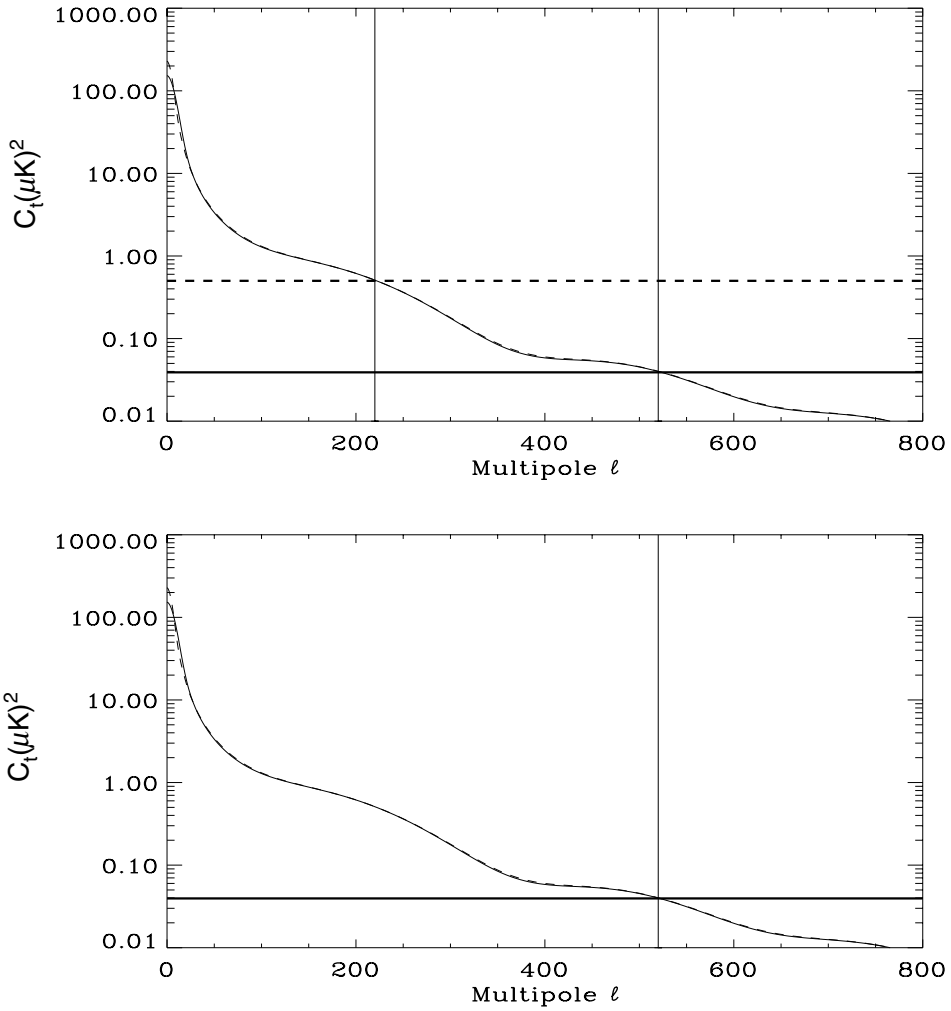

Figure 16. The plots show average signal and noise pseudo power spectra, plotted separately. The spectra are normalized so that they can be compared to the full-sky power spectrum. The solid and dashed curves, which almost overlap each other, are the signal pseudo power spectra for a 15° FWHM Gaussian Gabor window W_G and a corresponding top-hat window W_A respectively. In the upper plot the noise model shown in Fig. 16 was used. This noise model is increasing from the north pole and down to the edges of the patch. This is opposite of the Gaussian window and for this reason the Gaussian window gets higher signal to noise ratio. The solid horizontal line in the upper plot shows the noise pseudo power spectrum for the Gaussian window and the dashed horizontal line shows the noise pseudo power spectrum for the top-hat window. In the lower plot a uniform noise model was used so that the noise pseudo power spectra fall together and are shown as a solid vertical line. The figure shows how a Gabor window different from a top-hat can be used to increase the signal to noise ratio.

B. We used a standard CDM power spectrum and both patches *A* and *B* had a radius of 18° apodized with a 15° FWHM Gaussian Gabor window. In Fig. 24 we have plotted the diagonal of the normalized correlation matrix $(\langle\tilde{C}_\ell^A\tilde{C}_\ell^B\rangle - \langle\tilde{C}_\ell\rangle^2)/\langle\tilde{C}_\ell\rangle^2$. The angels θ we used were 6°, 12°, 24°, 30°, 36° and 180°. One sees clearly how the correlations drop with the distance. In the two last cases there were no common pixels in the patches. As one could expect, the correlations for the largest angels (the first few multipoles) do not drop that fast.

In Fig. 25 we have plotted two slices of the correlation matrix of $\tilde{C}_\ell$ for a single patch at $\ell = 400$ and $\ell = 800$. On the top we plotted the diagonals of the correlation matrices for separation angle $\theta = 30°$, $\theta = 36°$ and $\theta = 180°$. One sees that for the case where the patches do not have overlapping pixels, the whole diagonals have the same level as the far-off-diagonal elements in the $\theta = 0°$ matrix. When doing power spectrum estimation on one patch, the result did not change significantly when these far-off-diagonal elements were set to zero. For this reason one expects that when analysing several patches which do not overlap, simultaneously, the correlations between non-overlapping patches do not need to be taken into account. Note, however, that for the $\theta = 30°$ which means that there are only a few overlapping pixels, the approximation will not be that good as the level is orders of magnitude above the far-off-diagonals of the $\theta = 0°$ matrix. Another thing to note is that for the lowest multipoles, the correlation between patches is still high but we will also assume this part to be zero and attempt a joint analysis of non-overlapping patches.

The full correlation matrices for 0 and 30 degree separation are shown in Fig. 26. The figures show how the diagonal is dropping relative to the far off-diagonal elements.

In Fig. 27 the full correlation matrices for 36° and 180° separation is shown. For 36° separation one can see that the diagonal has almost disappeared with respect to the rest of the matrix whereas for 180° the diagonal has vanished completely. But the 'wall' at low multipoles remains.

In Fig. 28, we did a separate C_ℓ estimation on 146 non-overlapping patches with radius 18° apodized with a 15° FWHM Gaussian Gabor window. The patches where uniformly distributed over the sphere and uniform noise was added to the whole map. The figure shows

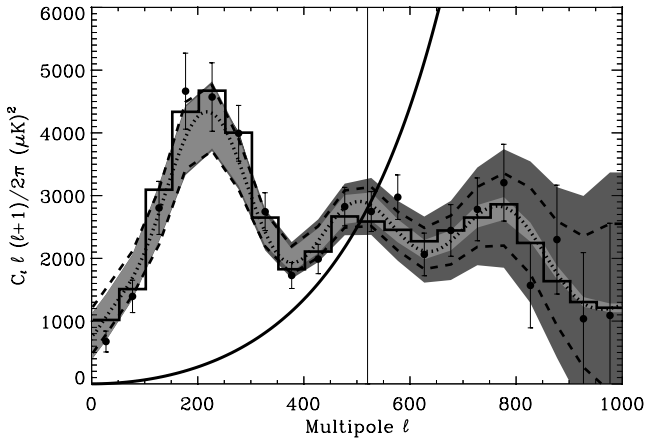

Figure 17. Same as Fig. 13 but for a standard CDM model. The noise is increasing from the centre and out to the edges while the Gaussian Gabor window has the opposite effect, giving an increased significance to pixels with less noise. As in Fig. 13, the shaded areas show the analytically calculated variance using the 'naive' formula for the uniform noise case. The dashed lines show the expected variance using the inverse of the Fisher matrix.

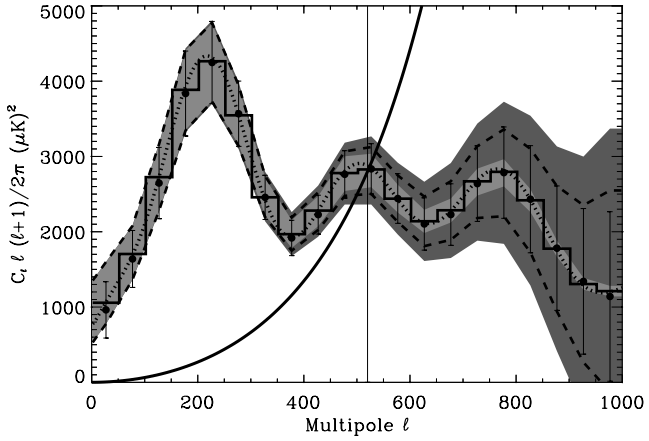

Figure 18. Same as Fig. 17 but with 5000 simulations and estimations averaged. The histogram now shows the binned average full-sky power spectrum. The error bars on the estimates are here the 1σ averages from Monte Carlo.

the average of the 146 C_ℓ estimates. One can see that the estimate seems to be approaching the full-sky power spectrum even at small multipoles.

Finally we made a joint analysis of all the 146 patches. The idea was to extend the data vector in the likelihood so that it contained the $N^{\rm in}$ $\tilde{C}_\ell$ from all the 146 patches. The data vector can then be written as $\boldsymbol{d} = \{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_{146}\}$ where $\boldsymbol{d}_i$ now denotes the whole data vector for patch number i. From the results above it seems to be a good approximation to assume that the correlation between $\tilde{C}_\ell$ from different patches is zero so that the correlation matrix will be block diagonal. Each block is then the correlation matrix for each individual patch. The log-likelihood can then simply be written as

$$L = \sum_{i=1}^{146} \boldsymbol{d}_i^{\rm T} \mathbf{M}_i^{-1} \boldsymbol{d}_i + \sum_{i=1}^{146} \ln\det \mathbf{M}_i, \tag{29}$$

where $\mathbf{M}_i$ is the correlation matrix for patch number i. In Fig. 29 the result of this joint analysis is shown. One can see that the full-sky power spectrum is well within the 2σ error bars of the estimates.

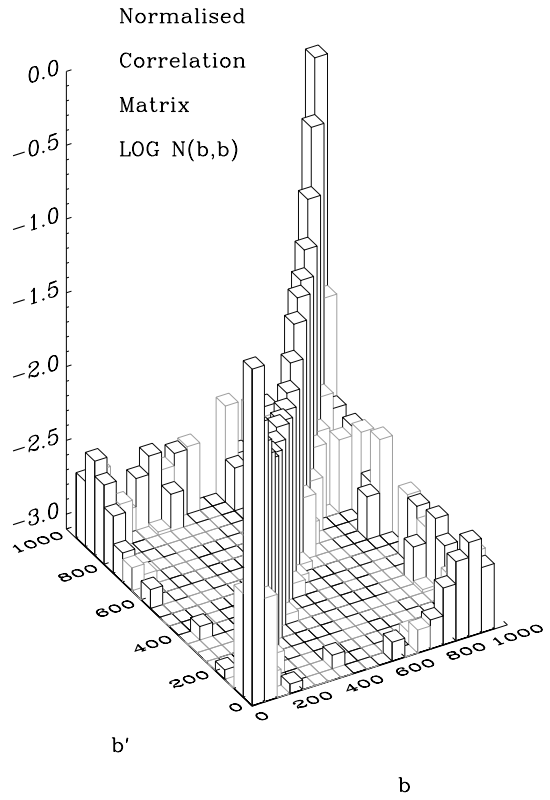

Figure 19. The average correlation matrix $N(b,b') \equiv \langle D_b D_{b'}\rangle / (\langle D_b\rangle\langle D_{b'}\rangle) - 1$ of the estimates in Fig. 18. The negative elements are shown in grey.

The method of combining patches on the sky for power spectrum analysis will be developed further in a forthcoming paper.

4.2 Monte Carlo simulations of the noise correlations and extension to correlated noise

The computation of the noise correlation matrix in the general case without any approximations takes $\sqrt{N_{\rm pix}} l_{\rm max}^2 (N^{\rm in})^2$ which is approximately $N_{\rm pix}^{3/2}(N^{\rm in})^2$ when $N_{\rm pix}$ is large ($N_{\rm pix} \approx \ell_{\rm max}^2$). When $N^{\rm in}$ is getting large this can be calculated quicker using Monte Carlo simulations (as was performed in Hivon et al. 2002). Finding the $\tilde{C}_\ell^{\rm N}$ from one noise map takes $\mathcal{O}(N_{\rm pix}^{3/2})$ operations, so using Monte Carlo simulations to find the whole noise matrix takes the $\mathcal{O}(N_{\rm pix}^{3/2} N_{\rm sim})$ operation where $N_{\rm sim}$ is the number of Monte Carlo simulations needed. Thus when $N_{\rm sim} \ll (N_{\rm in}^2)$ it will be advantageous using a Monte Carlo simulation if this gives the same result.

Also when the noise gets correlated, the analytic calculation of $\langle \tilde{a}_{\ell m}^{\rm N} \tilde{a}_{\ell' m}^{\rm N}\rangle$ will be very expensive. In this case another method for computing $\langle \tilde{a}_{\ell m}^{\rm N} \tilde{a}_{\ell' m}^{\rm N}\rangle$ will be necessary and Monte Carlo simulations could also prove useful. For a given noise model several noise realizations can be made and averaged to yield the noise correlation matrix and the $\langle \tilde{a}_{\ell m}^{\rm N} \tilde{a}_{\ell' m}^{\rm N}\rangle$ term needed in the estimation process. This is of course dependent on a method for fast evaluation of maps with different realizations.

In Fig. 30, the result of C_ℓ estimation with noise matrix and $\langle a_{\ell m}^{\rm N} a_{\ell' m}^{\rm N}\rangle$ computed with Monte Carlo is shown. Again a standard CDM power spectrum was used with a non-uniform white noise model and a 15° FWHM Gaussian Gabor window. In

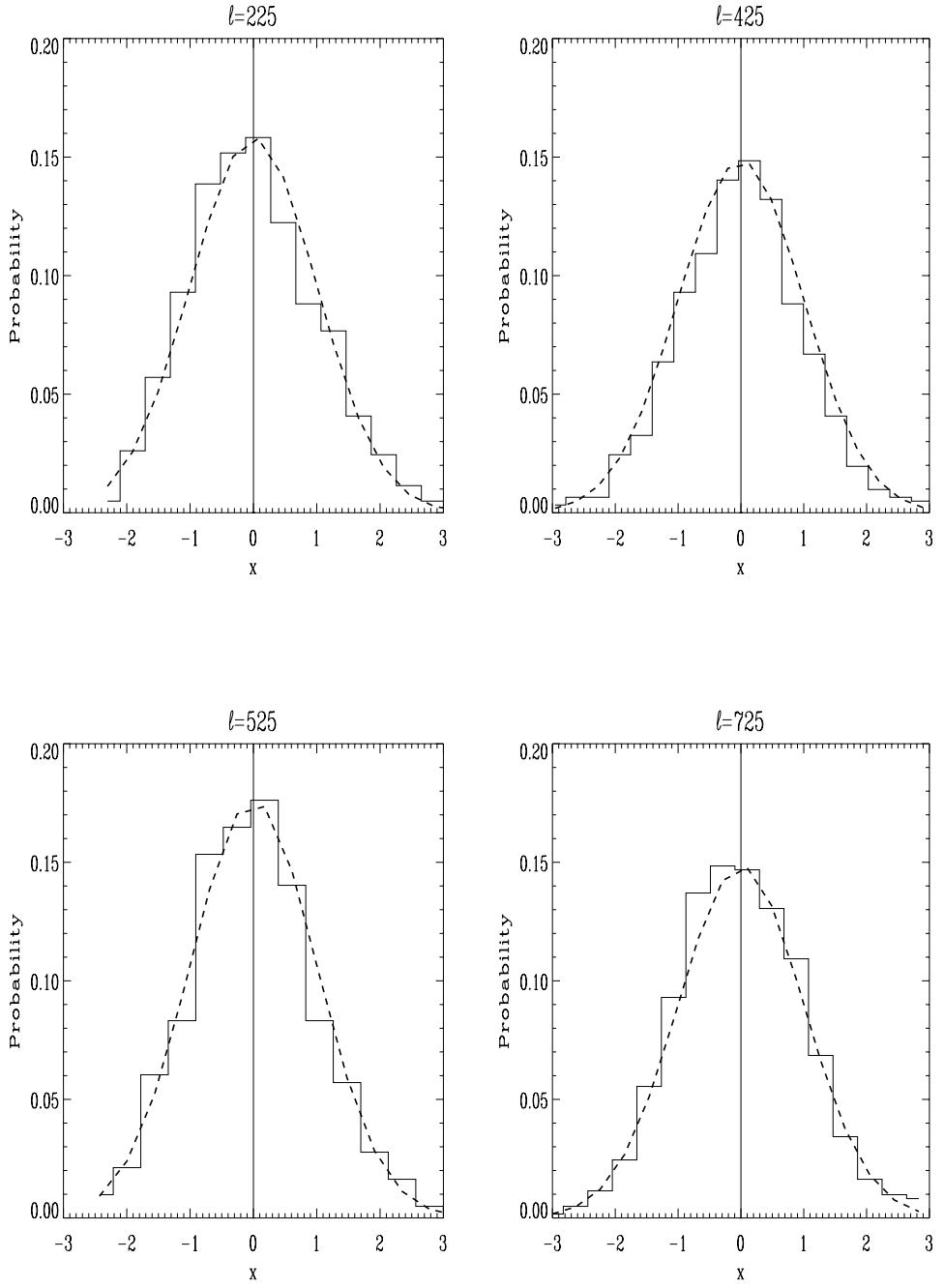

Figure 20. The probability distribution of the estimates D_b in Fig. 18. The variable x is given as $x = (D_b - \langle D_b \rangle)/\sqrt{\langle (D_b - \langle D_b \rangle)^2 \rangle}$. The dashed line is a Gaussian with mean and standard deviation taken from Monte Carlo. The plot shows bin estimates centred at $\ell = 225$, $\ell = 425$, $\ell = 525$ and $\ell = 725$.

the C_ℓ estimation $N^{\rm in} = 200\,\tilde{C}_\ell$ were used and $N^{\rm bin} = 20$ power spectrum bins were estimated. The noise matrices were calculated using (1) the analytical expression, (2) Monte Carlo with $N_{\rm sim} = 20\,000$ and (3) Monte Carlo with $N_{\rm sim} = 1000$. In Fig. 31, a slice of the correlation matrices for the different cases is shown for $\ell = 500$. The dashed line case (3) follows the solid line case (1) to a level of about 10^{-2} of the diagonal. The dotted line case (2) is roughly correct to about 10^{-1} times the value at the diagonal.

We did 100 estimations for each case and the average result is plotted in Fig. 30. The big dots are the results from case (1), the crosses on the right-hand side are the results from (2) and the crosses on the left-hand side the results from (3). The average estimates seem to be consistent, only in the highly noise dominated regime they start to deviate. For case (2), the error bars are for some multipoles higher and for some lower than the analytic case. The differences are at most 3 per cent. We conclude that using this many simulations, the error bars do not increase significantly over the analytic case. For case (3) the error bars are up to 17 per cent higher (and only higher) than the analytic case. It seems that 1000 simulations was not sufficient to keep the same accuracy of the estimates as when using analytic noise matrices. To keep the error bars close to the analytic case, it seems that a few thousand simulations are necessary.

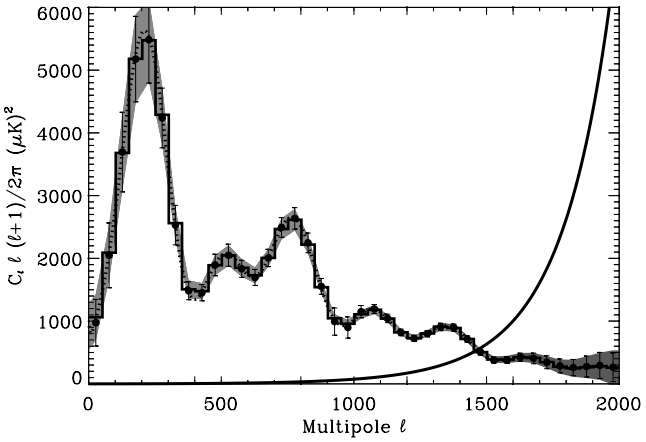

Figure 21. Same as Fig. 14 for 100 simulations where the beam and noise level was set according to the specifications of the *Planck* 143 GHz channel. Again, a 15° FWHM Gaussian Gabor window was used. The power spectrum was estimated in 40 bins between $\ell = 2$ and $\ell = 2048$.

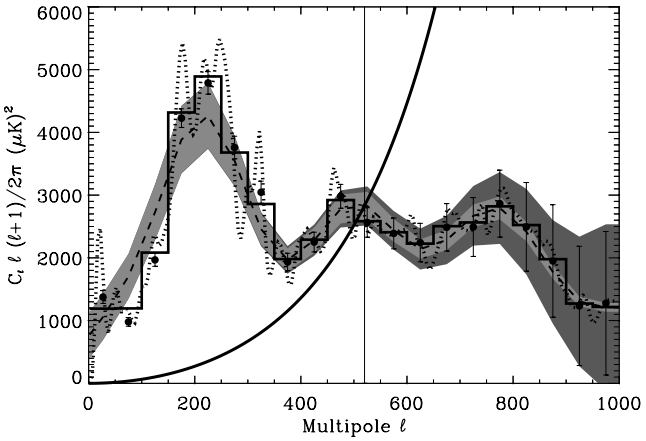

Figure 22. The average of 300 estimations where the input $\tilde{C}_\ell$ values were taken from simulations with a fixed CMB realization but varying noise realizations. The dotted line is the $N^{\rm in}$ input $\tilde{C}_\ell$ from the CMB realization without noise. The histogram is the same spectrum binned in $N^{\rm bin}$ bins. The dashed line is the binned average full-sky power spectrum from which this realization was made. The shaded areas around the binned full-sky spectrum show the variance with (dark) and without (light) noise. The solid line rising from the left to the right is the noise power spectrum.

5 CONCLUSION

In this paper, we propose to use the spherical harmonic transform of the sky apodized by a window function, or Gabor transform, as a fast and robust tool to estimate the CMB fluctuations power spectrum. It is known that the coupling between modes resulting from the analysis on a cut sky affects the shape of the measured pseudo power spectrum and the statistics of the C_ℓ coefficients (Wandelt et al. 2001; Hivon et al. 2002 and reference therein). In the case of axisymmetric windows we can compute analytically (in about $\ell^3_{\rm max}$ operations) the kernel relating the cut sky power spectrum to the full sky one for a Gaussian and top-hat profile we give an analytical relation between the spectral resolution attainable and the size of the sky window. Studying windows of different sizes, we show that for windows as small as 36° in radius, the measured power spectrum is undiscernible from the true one for ℓ larger than about 50.

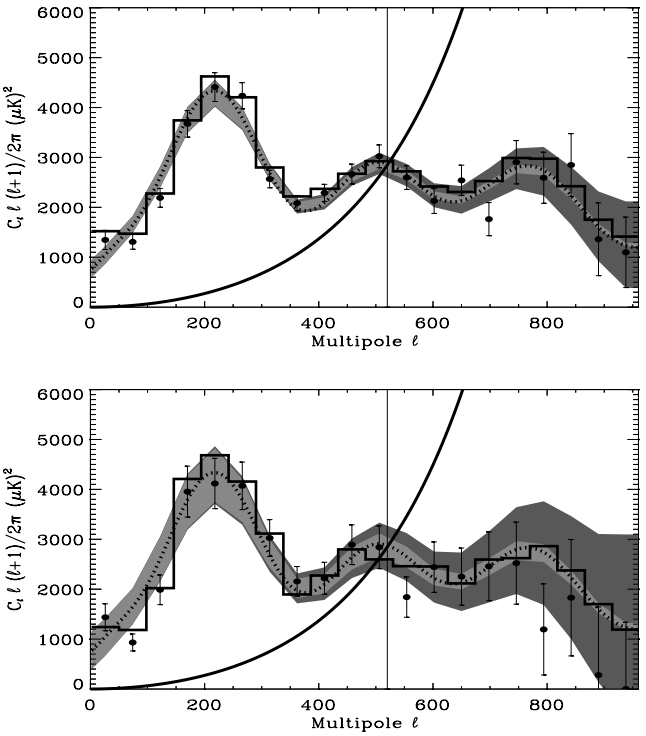

Figure 23. Estimates of $\tilde{C}_\ell$ using a Gaussian Gabor window (lower plot) and a top-hat window (upper plot). Here we used a uniform noise model with S/N = 1 at $\ell = 520$. The dotted line shows the average full-sky power spectrum and the histogram shows the input pseudo power spectrum without noise for this realization, binned in the same way as the estimates. The light shaded area shows the cosmic and sample variance around the binned average spectrum (not plotted). The dark shaded area has the variance due to noise included. The 1σ error bars on the estimates are taken from the inverse Fisher matrix. The solid line increasing from left to right is the noise power spectrum.

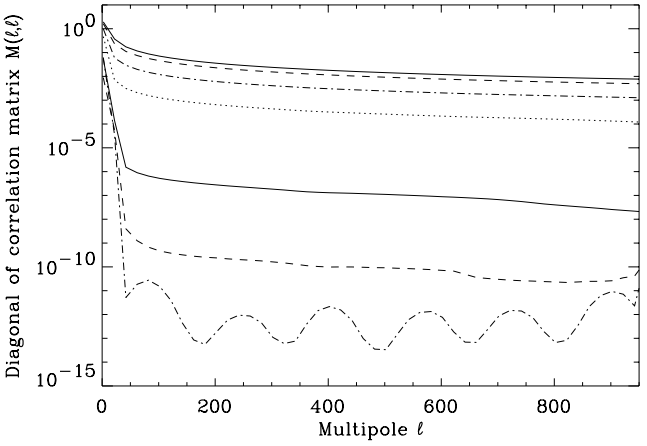

Figure 24. The correlation between $\tilde{C}_\ell$ between two patches A and B with an angular distance θ between the centres. A normal CDM power spectrum was used and the patches had an 18° radius apodized with a 15° FWHM Gaussian Gabor window. The figure shows the diagonal of the normalized correlation matrix $(\langle \tilde{C}^{\rm A}_\ell \tilde{C}^{\rm B}_\ell \rangle - \langle \tilde{C}_\ell \rangle^2)/\langle \tilde{C}_\ell \rangle^2$ where of course $\langle C_\ell \rangle = \langle C^{\rm A}_\ell \rangle = \langle C^{\rm B}_\ell \rangle$ The angles used are (from top to bottom on the figure) 0°, 6°, 12°, 18°, 30°, 36° and 180°.

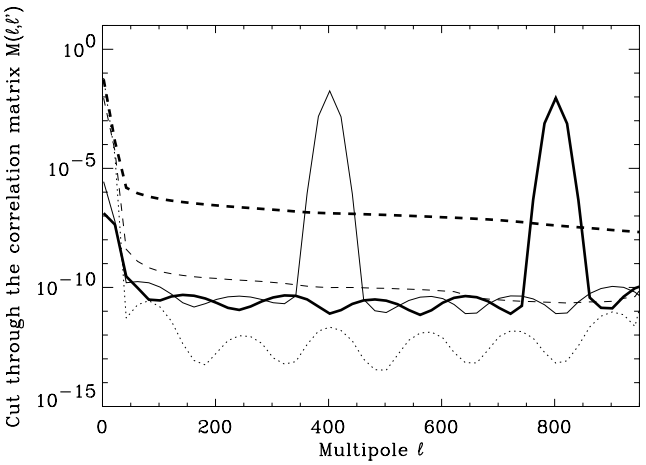

Figure 25. Slices of the correlation matrices which diagonals are shown in Fig. 24. The solid lines (thin and thick) show a slice of the $\theta = 0^\circ$ correlation matrix at $\ell = 400$ and $\ell = 800$, respectively. The dashed lines (thin and thick) show the diagonal of the correlation matrices for $\theta = 36^\circ$ and $\theta = 30^\circ$, respectively. The dotted line is the diagonal of the $\theta = 180^\circ$ matrix.

Noting that for large multipoles ($\ell \geqslant 100$ for windows with radii larger than 36°) the statistics of the pseudo-C_ℓ coefficients measured in Monte Carlo simulations is close to Gaussian, we suggest the use of the pseudo power spectrum as input data vector in a likelihood estimation. For the first time, we show how the correlation matrix between the pseudo power spectrum coefficients obtained on an axisymmetric window of arbitrary profile can be computed rapidly for any input power spectrum, based on a recurrence relation. The computation of the correlation matrix needs a precomputation (independent of the power spectrum) of $\ell_{\rm max} N_m (N^{\rm in})^2$ operations and each calculation of the correlation matrix with a given power spectrum takes $(N^{\rm in})^2 (N^{\rm bin})^2$ operations, where $N^{\rm in}$ is the number of input pseudo-C_ℓ coefficients used, $N^{\rm bin}$ is the number of estimated C_ℓ bins and N_m is a window-dependent factor ($N_m \approx 200$ for a Gaussian window and $\approx$400 for a top-hat window). The noise correlation matrix can also be computed by recurrence. For axisymmetric noise this is very quick ($N_{\rm pix} \ell_{\rm max}$ operations). For general non-uniform noise this takes some more time [between $(N^{\rm in})^2 \sqrt{N_{\rm pix}} \ell_{\rm max}$ and $(N_{\rm in})^2 \sqrt{N_{\rm pix}} \ell_{\rm max}^2$ operations dependent on the window profile and number of approximations]. For a Gaussian window with a sharply varying noise profile and a patch of sky similar in size to the one

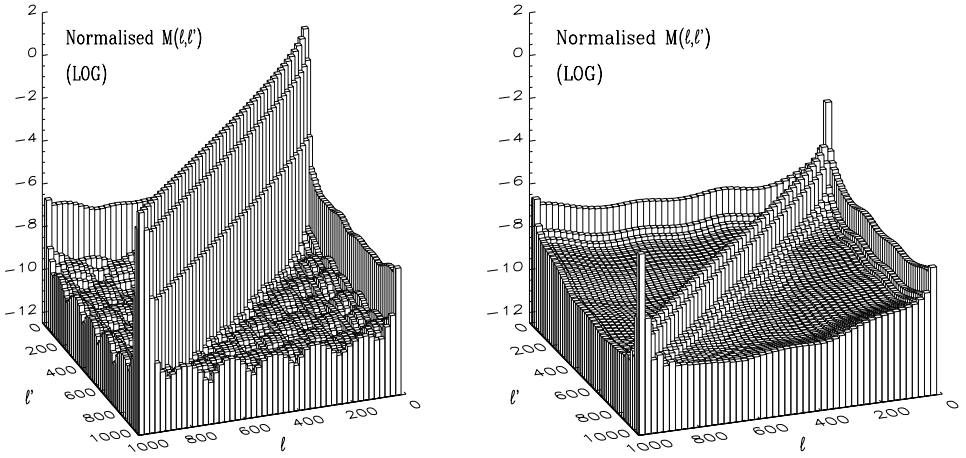

Figure 26. The figure shows the normalized correlation matrices $M(\ell, \ell') = (\langle C_\ell^{\rm A} C_\ell^{\rm B} \rangle - \langle C_\ell^{\rm A} \rangle \langle C_\ell^{\rm B} \rangle) / (\langle C_\ell^{\rm A} \rangle \langle C_\ell^{\rm N} \rangle)$ between pseudo-spectrum coefficients for two patches A and B of 18° radius and with 0° (left plot) and 30° separation. A Gaussian Gabor window with 15° FWHM was used. The aim of the plot is to show how correlations between $\tilde{C}_\ell$ from different patches drop when the distance between the two patches is about the FWHM of the Gaussian kernel.

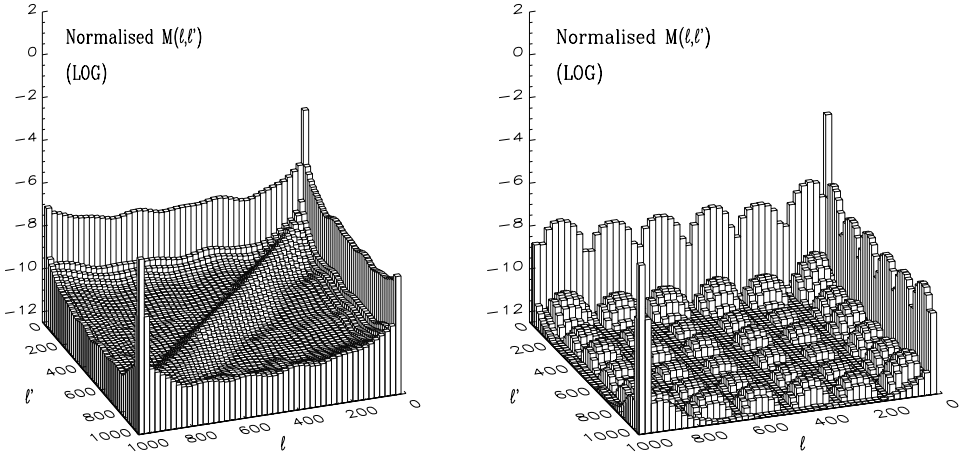

Figure 27. This figure shows the same as Fig. 26 but for 36° and 180° separation of the patches.

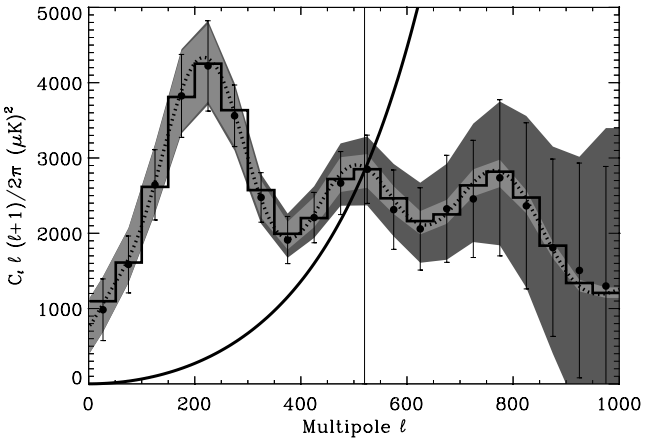

Figure 28. The average of 146 individual power spectrum estimations of 146 non-overlapping patches on the same CMB map with uniform noise added to it. The patches all had radius 18° degrees apodized with a 15° FWHM Gaussian Gabor window. The histogram shows the binned average of the 146 $\tilde{C}_\ell$ values from the different patches without noise. The dotted line is the average full-sky power spectrum and the shaded areas around the binned full-sky power spectrum (not plotted) show the theoretical 1σ variance with (dark) and without (light) noise. The solid line rising from the left to the right is the noise power spectrum.

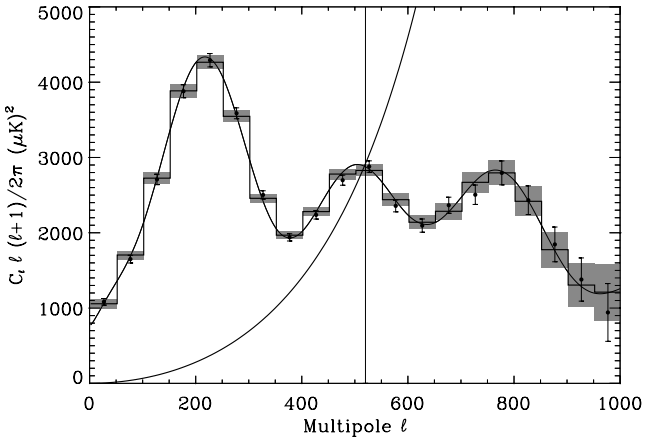

Figure 29. The result of a joint analysis of 146 patches on the same CMB map. The solid line shows the average full-sky power spectrum, the histogram shows the binned full-sky power spectrum and the shaded boxes show the expected 2σ deviations due to noise, cosmic and sample variance. The sizes of the shaded boxes were calculated from the formula for uniform noise (equation 26). The dots show the estimates with 2σ error bars taken from the Fisher matrix. As before the rising solid line is the noise power spectrum and the vertical line shows where S/N =1

observed by BOOMERANG (about 2 per cent of the sky) it takes about a day on one single 500-MHz processor. This is the computationally heaviest part of the method, but it only has to be done once. The inversion of the correlation matrix, which is the leading problem when doing likelihood analysis, is now overcome, as the size of the correlation matrix is so small that inversion is feasible. In the standard likelihood approach, the correlation matrix has dimensions $N_{\rm pix} \times N_{\rm pix}$ which needs $N_{\rm pix}^3$ operations to be inverted. In our approach, the size of the correlation matrix is $N^{\rm in} \times N^{\rm in}$, which in our example, where $N^{\rm in} = 200$, is inverted in a few seconds.

By doing Monte Carlo simulations of different experimental settings, we show that the likelihood estimator is unbiased. The error

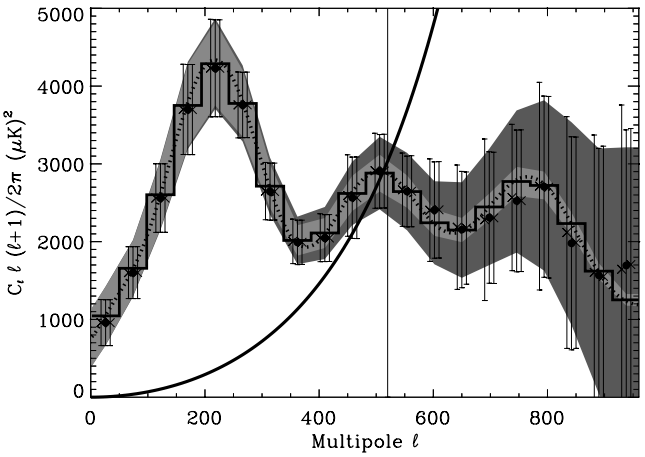

Figure 30. Same as Fig. 18 but with a different noise model. Here the average of 100 estimations is shown. The big dots in the middle of the bins show the result using analytical expressions for the noise matrices. The crosses on the left-hand side of each big dot show the results of using noise matrices from $N_{\rm sim} = 1000$ Monte Carlo simulations. The crosses on the right-hand side are for $N_{\rm sim} = 20\,000$ simulations.

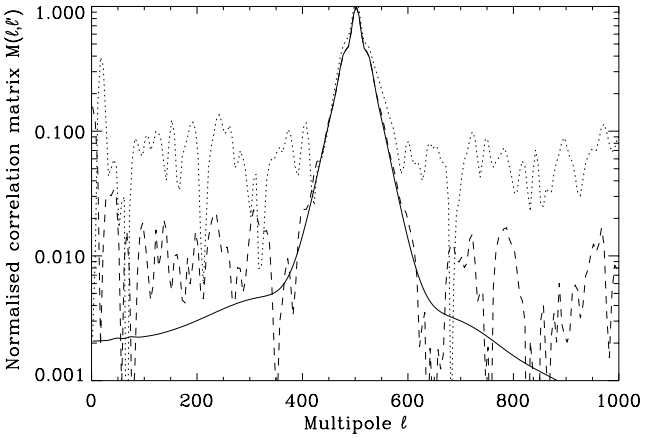

Figure 31. A slice of the noise correlation matrix at multipole $\ell = 500$. The correlation matrix was evaluated using the analytical formulae (solid line), 20 000 Monte Carlo simulations (dashed line) and 1000 Monte Carlo simulations (dotted line). The matrix is here normalized to be 1 at the diagonal.

bars were found using the inverse Fisher matrix and compared to the error bars obtained from Monte Carlo siumulations. There was an excellent agreement between the two sets of error bars. In (Hivon et al. 2002) it was shown that using a Gaussian apodization suppresses the signal such that the error bars on the estimated power spectrum become larger. In this paper we have shown that using a different window from the top-hat window can be important for increasing the signal-to-noise ratio in data with non-uniform noise. We applied a Gaussian window to an observed disc which had the noise level increasing from the centre of the disc and outwards similar to what one can expect around the ecliptic poles in scanning strategies like the ones of *MAP* and *Planck*. In this case the Gaussian window has a high value in the centre where signal-to-noise ratio per pixel is high and a low value close to the more noise dominated edges. We shown that for this noise profile using a Gaussian window increased the signal-to-noise ratio significantly over the top-hat window, showing that adapting the window to downweight noisy pixels gives better performance than a simple uniform weighting.

Finally, two extensions of the power spectrum estimation method were discussed. First it was shown that, for observed areas on the sky which are not axisymmetric, one can cover the area by several axisymmetric patches and make a joint analysis of the pseudo power spectrum coefficients from all the patches. Each of the patches can have a different window in order to optimize signal-to-noise ratio in each patch. This method will be extended in a forthcoming paper where we will discuss the use of the method for analysing *MAP* and *Planck* data sets. We also shown that the calculation of the noise correlation matrix can be quicker by Monte Carlo simulations if the number of pixels in the observed area is huge (about 10^6 pixels but dependent on the window shape). This may also be used in the case of correlated noise. We shown that a few thousand simulations are necessary to get the same accuracy in the power spectrum estimates as when using the analytic formula for the noise correlation matrix.

In Hansen & Górski (in preparation) we show that the power spectrum estimation method presented in this paper can easily be extended to polarization. By extending the data vector in the likelihood to also have the pseudo-C_ℓ from polarization, one can in a similar way estimate for the temperature and polarization power spectra jointly.

ACKNOWLEDGMENTS

We would like to thank A. J. Banday and B. D. Wandelt for helpful discussions. We acknowledge the use of HEALPIX (Górski, Hivon & Wandelt 1998) and CMBFAST (Seljak & Zaldariagga 1996). FKH was supported by a grant from the Norwegian Research Council.

REFERENCES

Balbi A., de Gasperis G., Natoli P., Vittorio N., 2002, preprint (astro-ph/0203159)
Bartlett J. G., 2000, AASS, 146, 506
Bersanelli M. et al., 1996, COBRAS/SAMBA: report on the phase A study. Available on-line at http://astro.estec.esa.nl/SA-general/Projects/Planck/
Bond J. R., 1995, Phys. Rev. Lett., 74, 4369
Bond J. R., Jaffe A. H., Knox L., 2000, ApJ, 533, 19
Borrill J., 1999, Phys. Rev. D, 59, 027302
De Bernardis P. et al., 2000, Nat, 404, 922
Dore O., Knox L., Peel A., 2001, Phys. Rev. D, 64, 083001
Durrer R., 2001, J. Phys. Stud., 5, 177
Gabor D., 1946, J. Inst. Elect. Eng., 93, 429
Górski K. M., Hivon E., Wandelt B. D., 1998, in Banday A. J., Sheth R. K., Da Costa L., eds, Analysis Issues for Large CMB Data Sets. ESO, Printpartners, Ipskamp, p. 37
Halverson N. W. et al., 2001, preprint (astro-ph/0104489)
Hanany S. et al., 2000, ApJ, 545, L5
Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, ApJ, 567, 2
Hobson M. P., Magueijo J., 1996, MNRAS, 283, 1133
Hu W., Sugiyama N., Silk J., 1997, Nat, 386, 37
Jaffe A. H. et al., 2001, Phys. Rev. Lett., 86, 3475
Jungman G., Kamionkowsky M., Kosowsky A., Spergel D. N., 1996, Phys. Rev. D, 54, 1332
Lee A. T. et al., 2001, ApJ, 561, L1
Maino D. et al., 2002, MNRAS, 334, 53
Netterfield C. B. et al., 2002, ApJ, 571, 604
Oh S. P., Spergel, D. N., Hinshaw G., 1999, ApJ, 510, 551
Pryke C., Halverson N. W., Leitch E. M., Kovac J., Carlstrom J. E., Holzapfel W. L., Dragovan M., 2002, ApJ, 568, 46
Seljak U., Zeldariagga M., 1996, ApJ, 1996, 469, 437
Smoot G. F. et al., 1992, ApJ, 396, L1
Stolyarov V., Hobson M. P., Ashdown M. A. J., Lasenby A. N., 2001, preprint (astro-ph/0105432)
Strohmer T., 1997, Proc. SampTA - Sampling Theory and Applications, p. 297
Szapudi I., Prunet S., Pogosyan D., Szalay A. S., Bond J. R., 2001, ApJ, 548, L115
Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
Wandelt B. D., Górski K. M., Hivon E. F., 2001, Phys. Rev. D, 64, 083003
Wandelt B. D., 2000, in Proc. MPA/MPE/ESO Conf., Mining the Sky. Preprint (astro-ph/0012416)
Wandelt B. D., Hansen F. K., 2001, preprint (astro-ph/0106515)

APPENDIX A: ROTATION MATRICES

A spherical function $T(\hat{\boldsymbol{n}})$ is rotated by the operator $\hat{D}(\alpha\beta\gamma)$ where $\alpha\beta\gamma$ are the three Euler angles for rotations (see Risbo 1996) and the inverse rotation is $\hat{D}(-\gamma-\beta-\alpha)$. For the spherical harmonic functions, this operator takes the form,

$$Y_{\ell m}(\hat{\boldsymbol{n}}') = \sum_{m'=-\ell}^{\ell} D^{\ell}_{m'm}(\alpha\beta\gamma) Y_{\ell m'}(\hat{\boldsymbol{n}}), \tag{A1}$$

where $D^{\ell}_{m'm}$ has the form

$$D^{\ell}_{m'm}(\alpha\beta\gamma) = \mathrm{e}^{\mathrm{i}m'\alpha} d^{\ell}_{m'm}(\beta) \mathrm{e}^{\mathrm{i}m\gamma}. \tag{A2}$$

Here $d^{\ell}_{m'm}(\beta)$ is a real coefficient with the following property:

$$d^{\ell}_{m'm}(\beta) = d^{\ell}_{mm'}(-\beta). \tag{A3}$$

The D-functions also have the following property:

$$D^{\ell}_{m'm}(\alpha\beta\gamma) = \sum_{m''} D^{\ell}_{m'm''}(\alpha_2\beta_2\gamma_2) D^{\ell}_{m''m}(\alpha_1\beta_1\gamma_1), \tag{A4}$$

where $(\alpha\beta\gamma)$ is the result of the two consecutive rotations $(\alpha_1\beta_1\gamma_1)$ and $(\alpha_2\beta_2\gamma_2)$.

The complex conjugate of the rotation matrices can be written as

$$D^{\ell *}_{mm'} = (-1)^{m+m'} D^{\ell}_{(-m)(-m')}. \tag{A5}$$

APPENDIX B: SOME WIGNER SYMBOL RELATIONS

Throughout the paper, the Wigner 3j Symbols will be used frequently. Here are some relations for these symbols, which are used. The orthogonality relation is,

$$\sum_{mm'} \begin{pmatrix} \ell & \ell' & \ell'' \\ m & m' & m'' \end{pmatrix} \begin{pmatrix} \ell & \ell' & L'' \\ m & m' & M'' \end{pmatrix} = (2\ell''+1)^{-1} \delta_{\ell'' L''} \delta_{m'' M''}. \tag{B1}$$

The Wigner 3j Symbols can be represented as an integral of rotation matrices (see Appendix A),

$$\frac{1}{8\pi^2} \int \mathrm{d}\cos\theta \, \mathrm{d}\phi \, \mathrm{d}\gamma \, D^{\ell}_{m_1 m_1'} D^{\ell'}_{m_2 m_2'} D^{\ell''}_{m_3 m_3'} = \begin{pmatrix} \ell & \ell' & \ell'' \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ m_1' & m_2' & m_3' \end{pmatrix}. \tag{B2}$$

This expression can be reduced to,

$$\int d\hat{\boldsymbol{n}} Y_{\ell m}(\hat{\boldsymbol{n}}) Y_{\ell' m'}(\hat{\boldsymbol{n}}) Y_{\ell'' m''}(\hat{\boldsymbol{n}}) = \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \times \begin{pmatrix} \ell & \ell' & \ell'' \\ m & m' & m'' \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}. \tag{B3}$$

APPENDIX C: RECURRENCE RELATION

It is important for the precalculations to the likelihood analysis that the calculation of $h(\ell, \ell', m)$ is fast. For this reason a recurrence relation for $h(\ell, \ell', m)$ would be helpful. To speed up the calculation of the noise correlation matrix for non-axisymmetric noise, it would also help if one had a more general recurrence relation for $h(\ell, \ell', m, m')$. We will now show how to find such a recurrence for these objects which we now call $A^{m'm}_{\ell'\ell}$ to simplify notation (and for the notation to comply with Wandelt et al. 2001). The definition is, again,

$$A^{m'm}_{\ell'\ell} = \int_a^b G(\hat{\boldsymbol{n}}) Y^*_{\ell m}(\hat{\boldsymbol{n}}) Y_{\ell' m'}(\hat{\boldsymbol{n}})\, \mathrm{d}\hat{\boldsymbol{n}}, \tag{C1}$$

where $G(\hat{\boldsymbol{n}}) = G(\theta, \phi)$ is a general function and $Y_{\ell m}$ are the spherical harmonics which can be factorized into one part dependent on θ and one part dependent on ϕ in the following way,

$$Y_{\ell m}(\theta, \phi) = \lambda_{\ell m}(\cos\theta) \mathrm{e}^{-\mathrm{i}\phi m}. \tag{C2}$$

Now one can write

$$A^{m'm}_{\ell'\ell} = \int \mathrm{d}\cos\theta \lambda_{\ell m}(\cos\theta)\lambda_{\ell' m'}(\cos\theta) \int \mathrm{d}\phi \mathrm{e}^{-\mathrm{i}\phi(m-m')} G(\theta, \phi) \tag{C3}$$

$$\equiv \int \mathrm{d}\cos\theta \lambda_{\ell m}(\cos\theta)\lambda_{\ell' m'}(\cos\theta) F_{m'm}(\theta), \tag{C4}$$

where $F_{m'm}(\theta)$ is simply the Fourier transform of the window at each θ. The quantities $A^{m'm}_{\ell'\ell}$ and $F_{m'm}(\theta)$ are in general complex quantities obeying,

$$A^{m'm}_{\ell'\ell} = (A^{mm'}_{\ell\ell'})^* F_{m'm}(\theta) = (F_{mm'}(\theta))^* \tag{C5}$$

The $A^{m'm}_{\ell'\ell}$ can be expressed as

$$A^{m'm}_{\ell'\ell} = \frac{\sqrt{(2\ell'+1)(2\ell+1)}}{2} \sqrt{\frac{(\ell'-m')!(\ell-m)!}{(\ell'+m')!(\ell+m)!}} I^{m'm}_{\ell'\ell}, \tag{C6}$$

where $I^{m'm}_{\ell'\ell}$ is defined as:

$$I^{m'm}_{\ell'\ell} = \int_a^b F_{m'm}(x) P^m_\ell(x) P^{m'}_{\ell'}(x)\, \mathrm{d}x. \tag{C7}$$

The following relation for the Legendre Polynomials will be used:

$$x P^m_\ell = \frac{\ell-m+1}{2\ell+1} P^m_{\ell+1} + \frac{\ell+m}{2\ell+1} P^m_{\ell-1}. \tag{C8}$$

We now define the object $X^{m'm}_{\ell'\ell}$ as

$$X^{m'm}_{\ell'\ell} = \int_a^b F_{m'm}(x) x P^m_\ell P^{m'}_{\ell'}\, dx. \tag{C9}$$

Using relation (C8) in this definition, one gets

$$X^{m'm}_{\ell'\ell} = \frac{\ell-m+1}{2\ell+1} I^{m'm}_{\ell'(\ell+1)} + \frac{\ell+m}{2\ell+1} I^{m'm}_{\ell'(\ell-1)}. \tag{C10}$$

One can also exchange (ℓ, ℓ') and (m, m') to get

$$X^{mm'}_{\ell\ell'} = \frac{\ell'-m'+1}{2\ell'+1} I^{mm'}_{\ell(\ell'+1)} + \frac{\ell'+m'}{2\ell'+1} I^{mm'}_{\ell(\ell'-1)}. \tag{C11}$$

Taking the complex conjugate of the first expression and subtracting the last, one has

$$\left(X^{m'm}_{\ell'\ell}\right)^* - X^{mm'}_{\ell\ell'} = 0$$

$$= \frac{\ell-m+1}{2\ell+1} \left[I^{m'm}_{\ell'(\ell+1)}\right]^* + \frac{\ell+m}{2\ell+1} \left[I^{m'm}_{\ell'(\ell-1)}\right]^* \tag{C12}$$

$$- \frac{\ell'-m+1}{2\ell'+1} I^{mm'}_{\ell(\ell'+1)} - \frac{\ell'+m}{2\ell'+1} I^{mm'}_{\ell(\ell'-1)} \tag{C13}$$

Then setting $\ell' = \ell' - 1$ one gets:

$$I^{m'm}_{\ell'\ell} = \frac{2\ell'-1}{\ell'-m'} \left(\frac{\ell-m+1}{2\ell+1} I^{m'm}_{(\ell'-1)(\ell+1)} + \frac{\ell+m}{2\ell+1} I^{m'm}_{(\ell'-1)(\ell-1)} - \frac{\ell'+m-1}{2\ell'-1} I^{m'm}_{(\ell'-2)\ell} \right) \tag{C14}$$

Using equation (C6), one can express this as

$$A^{m'm}_{\ell'\ell} = \frac{1}{\sqrt{\ell'^2 - m'^2}} \left\{ \sqrt{\frac{(4\ell'^2-1)[(\ell+1)^2-m^2]}{(2\ell+1)(2\ell+3)}} A^{m'm}_{(\ell'-1)(\ell+1)} + \sqrt{\frac{(4\ell'^2-1)(\ell^2-m^2)}{4\ell^2-1}} A^{m'm}_{(\ell'-1)(\ell-1)} - \sqrt{\frac{(2\ell'+1)[(\ell'-1)^2-m'^2]}{2\ell'-3}} A^{m'm}_{(\ell'-2)\ell} \right\}, \tag{C15}$$

which is the final recurrence relation. The $A^{m'm}_{m'\ell}$ elements must be provided before the recurrence is started. Then for each (m, m'), set $\ell' = m' + 1$ and let ℓ go from ℓ' and upwards, then set $\ell' = m' + 2$ and again let ℓ go from ℓ' and upwards. Continue to the desired size of ℓ'. Note that, in order to get all objects up to $A^{m'm}_{\ell_{\max}\ell_{\max}}$, one always needs to go up to $\ell = 2\ell_{\max}$ during recursion. This is because of the $A^{m'm}_{(\ell'-1)(\ell+1)}$ term which demands an object indexed $(\ell+1)$ in the previous ℓ' row.

To start the recurrence, one can precompute the $A^{m'm}_{m'\ell}$ factors quickly and easily using Fast Fourier Transform (FFT) and a sum over rings on the grid. For example, for the HEALPIX grid, we did it the following way,

$$A^{m'm}_{m'\ell} = \sum_r \lambda^r_{m'm'} \lambda^r_{\ell m} \sum_{j=0}^{N_r-1} \mathrm{e}^{-2\pi\mathrm{i}j/N_r(m-m')} G_{rj}, \tag{C16}$$

where the last part is the Fourier transform of the Gabor window, calculated by FFT, r is ring number on the grid and j is azimuthal position on each ring. Ring r has N_r pixels.

It turns out that the recurrence can be numerically unstable, dependent on the window and multipole, and in order to avoid problems we (using double precision numbers) restart the recurrence with a new set of precomputed $A^{m'm}_{\ell'\ell}$ for every 50th ℓ' row. However, for most of the windows and multipoles that we tested the recurrence can run for hundreds of ℓ-rows without problems.

APPENDIX D: ROTATIONAL INVARIANCE

It was shown that the average $\langle\tilde{C}_\ell\rangle$ is invariant under rotations of the Gabor window. We will now show that the non-averaged $\tilde{C}_\ell$ are rotationally invariant under any rotation of the sky *and* Gabor window by the same angle. This fact justifies that we can always put the window on the north pole as this simplifies the calculations. In the following we will use the rotation matrices $D^\ell_{mm'}$ described in Appendix (A). Consider a rotation of the sky and window by the angles $(-\gamma - \beta - \alpha)$. Then the $\tilde{a}_{\ell m}$ becomes,

$$\tilde{a}^{\rm rot}_{\ell m} = \int {\rm d}\hat{\boldsymbol{n}}[\hat{D}(-\gamma-\beta-\alpha)T(\hat{\boldsymbol{n}})G(\hat{\boldsymbol{n}})]Y^*_{\ell m}(\hat{\boldsymbol{n}}). \quad \text{(D1)}$$

If one makes the inverse rotation of the integration angle $\hat{\boldsymbol{n}}$, one can write this as

$$\tilde{a}^{\rm rot}_{\ell m} = \int {\rm d}\hat{\boldsymbol{n}} T(\hat{\boldsymbol{n}})G(\hat{\boldsymbol{n}}) \left[\hat{D}^*(\alpha\beta\gamma)Y^*_{\ell m}(\hat{\boldsymbol{n}})\right], \quad \text{(D2)}$$

which is just

$$\tilde{a}^{\rm rot}_{\ell m} = \sum_{m'} D^{\ell *}_{m'm}(\alpha\beta\gamma) \int {\rm d}\hat{\boldsymbol{n}} T(\hat{\boldsymbol{n}})G(\hat{\boldsymbol{n}})Y^*_{\ell m'}(\hat{\boldsymbol{n}}). \quad \text{(D3)}$$

One can identify the last integral as the normal $\tilde{a}_{\ell m}$.

$$\tilde{a}^{\rm rot}_{\ell m} = \sum_{m'} D^{\ell *}_{m'm}(\alpha\beta\gamma)\tilde{a}_{\ell m}. \quad \text{(D4)}$$

Thus the $\tilde{a}_{\ell m}$ are *not* rotationally invariant. Rotation mixes m-modes for a given ℓ-value.

Now to the $\tilde{C}_l$. One has that

$$\tilde{C}^{\rm rot}_\ell = \frac{1}{2\ell+1}\sum_m a^{\rm rot}_{\ell m} a^{\rm rot *}_{\ell m} \quad \text{(D5)}$$

$$= \frac{1}{2\ell+1}\sum_m\sum_{m'}\sum_{m''} D^\ell_{m'm}(\alpha\beta\gamma)D^{\ell *}_{m''m}(\alpha\beta\gamma)\tilde{a}_{\ell m'}\tilde{a}^*_{\ell m''} \quad \text{(D6)}$$

$$= \frac{1}{2\ell+1}\sum_{m'm''}\tilde{a}_{\ell m'}\tilde{a}^*_{\ell m''}\sum_m D^\ell_{m'm}(\alpha\beta\gamma)D^{\ell *}_{m''m}(\alpha\beta\gamma). \quad \text{(D7)}$$

Using the properties given in Appendix (A), one can write the last D-function on the last line as,

$$\hat{D}^{\ell *}_{m''m}(\alpha\beta\gamma) = D^\ell_{m''m}(-\alpha\beta-\gamma) \quad \text{(D8)}$$

$$= {\rm e}^{-{\rm i}m''\alpha} d^\ell_{m''m}(\beta){\rm e}^{-{\rm i}m\gamma} \quad \text{(D9)}$$

$$= {\rm e}^{-{\rm i}m''\alpha} d^\ell_{mm''}(-\beta){\rm e}^{-{\rm i}m\gamma} \quad \text{(D10)}$$

$$= D^\ell_{mm''}(-\gamma-\beta-\alpha). \quad \text{(D11)}$$

Knowing that $(-\gamma-\beta-\alpha)$ is the inverse rotation of $(\alpha\beta\gamma)$ one can write,

$$\sum_m D^\ell_{m'm}(\alpha\beta\gamma)D^{\ell *}_{m''m}(\alpha\beta\gamma)$$

$$= \sum_m D^\ell_{m'm}(\alpha\beta\gamma)D^\ell_{mm''}(-\gamma-\beta-\alpha) \quad \text{(D12)}$$

$$= D^\ell_{m'm''}(000) = \delta_{m'm''} \quad \text{(D13)}$$

So one gets,

$$\tilde{C}^{\rm rot}_\ell = \tilde{C}_\ell. \quad \text{(D14)}$$

APPENDIX E: THE CORRELATION MATRIX

To do fast likelihood analysis with $\tilde{C}_\ell$ one needs to be able to calculate $\langle\tilde{C}_\ell\rangle$ and the correlations $\langle\tilde{C}_\ell\tilde{C}_{\ell'}\rangle$ fast. Calculating the average $\langle\tilde{C}_\ell\rangle$ by formula (10) using the analytic expression (11) for the kernel is not very fast. It turns out that a faster way of evaluating the kernel is by using direct integration (summation on the pixelized sphere) and then, as shown in Appendix C, recurrence. By means of an integral, one can then write the $\tilde{a}_{\ell m}$ as (now assuming that $\hat{\boldsymbol{n}}_0$ is on the north pole),

$$\tilde{a}_{\ell m} = \sum_{\ell' m'} a_{\ell' m'} \int G(\hat{\boldsymbol{n}})Y^*_{\ell m}(\hat{\boldsymbol{n}})Y_{\ell' m'}(\hat{\boldsymbol{n}})\,{\rm d}\hat{\boldsymbol{n}}$$

$$= \sum_{\ell' m'} a_{\ell' m'} \int G(\theta)\lambda_{\ell m}(\theta)\lambda_{\ell' m'}(\theta){\rm d}\cos\theta \underbrace{\int {\rm e}^{-{\rm i}\phi(m-m')}\,{\rm d}\phi}_{2\pi\delta_{mm'}}$$

$$= \sum_{\ell'} a_{\ell' m} 2\pi \int G(\theta)\lambda_{\ell m}(\theta)\lambda_{\ell' m}(\theta)\,{\rm d}\cos\theta$$

$$\equiv \sum_{\ell'} a_{\ell' m} h(\ell, \ell', m), \quad \text{(E1)}$$

where the last line defines $h(\ell, \ell', m)$ and $\lambda_{\ell m}(\theta)$ is given by,

$$Y_{\ell m}(\theta,\phi) = \lambda_{\ell m}(\theta){\rm e}^{-{\rm i}\phi m}. \quad \text{(E2)}$$

Using this form, one gets,

$$\langle\tilde{C}_\ell\rangle = \frac{1}{2\ell+1}\sum_{\ell'} C_{\ell'}\sum_m h^2(\ell,\ell',m). \quad \text{(E3)}$$

To obtain this expression, $\hat{\boldsymbol{n}}_0$ was on the north pole, but as was shown, the $\langle\tilde{C}_\ell\rangle$s are rotationally invariant, that is $\langle\tilde{C}_\ell\rangle$ remains the same if one rotates the Gabor window so that it is centred on the north pole.

When using real CMB data, the observed temperature map is always pixelized. Thus an integral over the sphere has to be replaced by a sum over pixels. In this case, the formula for $h(\ell, \ell', m)$ has to be replaced by

$$h(\ell,\ell',m) = \sum_j G_j \lambda^j_{\ell m}\lambda^j_{\ell' m}\Delta_j, \quad \text{(E4)}$$

where the index j is the pixel number replacing the angle θ and Δ_j is the area of pixel j. Using a pixelization scheme like HEALPIX (Górski et al. 1998), which has a structure of azimuthal rings going from the north to the south pole with N_r pixels in ring r and equal area for each pixel $\Delta_j = \Delta$, this can be written as

$$h(\ell,\ell',m) = \Delta\sum_r\sum_{p=0}^{N_r-1} G_r\lambda^r_{\ell m}\lambda^r_{\ell' m},$$

$$= \Delta\sum_r N_r G_r\lambda^r_{\ell m}\lambda^r_{\ell' m}. \quad \text{(E5)}$$

Here the sum over pixels is split into a sum over rings r and a sum over the pixels in each ring p. The first sum goes over all rings which have $\theta < \theta_{\rm C}$.

Using this expression for the $\tilde{a}_{\ell m}$ one can now find the correlation matrix

$$\langle\tilde{C}_\ell\tilde{C}_{\ell'}\rangle = \sum_{mm'}\frac{\langle\tilde{a}^*_{\ell m}\tilde{a}_{\ell m}\tilde{a}^*_{\ell' m'}\tilde{a}_{\ell' m'}\rangle}{(2\ell+1)(2\ell'+1)}. \quad \text{(E6)}$$

In this expression one can use relation (E1) to get,

$$\langle\tilde{C}_\ell\tilde{C}_{\ell'}\rangle = \frac{1}{(2\ell+1)(2\ell'+1)}\sum_{mm'}\sum_{LL'KK'}\langle a^*_{Lm}a_{L'm}a^*_{Km'}a_{K'm'}\rangle$$
$$\times h(\ell,L,m)h(\ell,L',m)h(\ell',K,m')h(\ell',K',m')$$
$$= \frac{1}{(2\ell+1)(2\ell'+1)}\sum_{mm'}\sum_{LL'KK'}\left[\langle a^*_{Lm}a_{L'm}\rangle\langle a^*_{Km'}a_{K'm'}\rangle\right.$$
$$\left. + \langle a^*_{Lm}a^*_{Km'}\rangle\langle a_{L'm}a_{K'm'}\rangle + \langle a^*_{Lm}a_{K'm'}\rangle\langle a_{L'm}a^*_{Km'}\rangle\right]$$
$$\times h(\ell,L,m)h(\ell,L',m)h(\ell',K,m')h(\ell',K',m'). \quad \text{(E7)}$$

Clearly the first term is just the product $\langle\tilde{C}_\ell\rangle\langle\tilde{C}_{\ell'}\rangle$, and the two last terms are equal [using $a^*_{Km'} = a_{Km'}(-1)^{m'}$ and $a_{K'm'} = (-1)^{m'}a^*_{K'm'}$] so one gets,

$$M_{ij} = \frac{2}{(2\ell+1)(2\ell'+1)}\sum_m\left[\sum_L C_L h(\ell_i, L, m)h(\ell_j, L, m)\right]^2. \tag{E8}$$

This is one of the main results of this paper, as the formula allows one to analytically calculate the correlation matrix needed for likelihood analysis. Another main result is the recurrence deduced in Appendix C which allows fast evaluation of the $h(\ell, \ell', m)$ functions and thereby this correlation matrix.

By using the binning of the power spectrum described in equation (22), the correlation matrix can be calculated faster if it is written as

$$M^{\rm S}_{ij} = \sum_b\sum_{b'} D_b D_{b'}\chi(b, b', i, j), \tag{E9}$$

where $\chi(b, b', i, j)$ is given as,

$$\chi(b, b', i, j) \equiv \frac{2}{(2\ell_i+1)(2\ell_j+1)} \times\sum_m\left[\sum_{l\epsilon b} B^2_\ell\ell(\ell+1)h(\ell, \ell_i, m)h(\ell, \ell_j, m)\right] \times\left[\sum_{l\epsilon b'} B^2_\ell\ell(\ell+1)h(\ell, \ell_i, m)h(\ell, \ell_j, m)\right], \tag{E10}$$

which is precomputed. The sums over ℓ here go over the ℓ values in each specific bin b. One sees that computing the likelihood takes of the order $(N^{\rm bin})^2(N^{\rm in})^2$ operations whereas the precomputation of the factor $\chi(b, b', i, j)$ goes as $\ell_{\rm max}(N^{\rm in})^2 N_m$ where N_m is the number of m values used. Note that the multipole coefficients of the beam B_ℓ are also included. The reason is that the input data is always affected by the beam and this is corrected for by using the beam convolved full-sky power spectrum $C_\ell B^2_\ell$.

The sum over m in the expressions for the covariance matrix and $\langle\tilde{C}_\ell\rangle$ can be limited. The h-functions are rapidly decreasing for increasing m for Gaussian and top-hat windows. For Gaussian Gabor windows it seems that one can cut the sums over m at $N_m = 200$ to high accuracy. For top-hat windows, the sum should be extended to $N_m = 400$.

APPENDIX F: INCLUDING WHITE NOISE

In this appendix the total $\langle\tilde{C}_\ell\rangle$ and the correlation matrix including contributions from white noise is found analytically. We assume that each pixel j has a noise temperature denoted by n_j, with the following properties,

$$\langle n_j\rangle = 0, \langle n_j n_{j'}\rangle = \delta_{jj'}\sigma^2_j, \tag{F1}$$

where σ_j is the noise variance in pixel j. Then one has the following expressions for the $a_{\ell m}$ and C_l (we use superscript N for noise quantities),

$$a^{\rm N}_{\ell m} = \sum_j n_j Y^{j*}_{\ell m}\Delta_j \tag{F2}$$

$$\left\langle a^{\rm N}_{\ell m}{a^{\rm N}_{\ell' m'}}^*\right\rangle = \sum_{jj'}\left\langle n_j n_{j'}\right\rangle Y^j_{\ell m}{Y^{j'}_{\ell m}}^*\Delta_j\Delta_{j'} = \sum_j\sigma^2_j Y^j_{\ell m}{Y^j_{\ell' m'}}^*\Delta^2_j \tag{F3}$$

$$\left\langle C^{\rm N}_\ell\right\rangle = \frac{1}{2\ell+1}\sum_m\left\langle a^{\rm N}_{\ell m}{a^{\rm N}_{\ell m}}^*\right\rangle = \frac{1}{4\pi}\sum_j\sigma^2_j\Delta^2_j. \tag{F4}$$

Here $Y^j_{\ell m}$ is the spherical harmonic of the pixel centre of pixel j. Similarly, for the windowed coefficients, one gets,

$$\tilde{a}^{\rm N}_{\ell m} = \sum_j G_j n_j\Delta_j Y^{j*}_{\ell m} \tag{F5}$$

$$\tilde{C}^{\rm N}_\ell = \frac{1}{4\pi}\sum_j\sigma^2_j G^2_j\Delta^2_j \tag{F6}$$

The next step is to find the noise correlation matrix,

$$\left\langle\tilde{C}^{\rm N}_\ell\tilde{C}^{\rm N}_{\ell'}\right\rangle = \frac{1}{(2\ell+1)(2\ell'+1)}\sum_{mm'}\sum_{jj'kk'}\Delta_j\Delta_{j'}\Delta_k\Delta_{k'} \tag{F7}$$

$$\times\langle n_j n_{j'} n_k n_{k'}\rangle G_j G_{j'} G_k G_{k'} Y^j_{\ell m} Y^{j'}_{\ell m} Y^k_{\ell' m'} Y^{k'}_{\ell' m'} \tag{F8}$$

$$= \left\langle C^{\rm N}_\ell\right\rangle\left\langle C^{\rm N}_{\ell'}\right\rangle + M^{\rm N}_{\ell\ell'}, \tag{F9}$$

where $M^{\rm N}_{\ell\ell'}$ can be written as,

$$M^{\rm N}_{\ell\ell'} = \frac{2}{(2\ell+1)(2\ell'+1)}\sum_{mm'}\left(\sum_j\Delta^2_j G^2_j\sigma^2_j Y^j_{\ell m}{Y^j_{\ell' m'}}^*\right)^2 \tag{F10}$$

For pixelization schemes like HEALPIX, the expression can be evaluated fast using FFT. This is apparent when one writes the sum over pixels as a double sum over rings and pixels per ring.

$$\sum_j\Delta^2_j G^2_j\sigma^2_j Y^j_{\ell m}{Y^j_{\ell' m'}}^* = \sum_r\Lambda^r_{\ell m}\Lambda^r_{\ell' m'} \times\sum_{p=0}^{N_r-1}{\rm e}^{-{\rm i}\phi_p(m-m')}\Delta^2 G^2_r\sigma^2_{r,p}. \tag{F11}$$

In the case of an axisymmetric noise model, this expression becomes even easier, which is apparent when writing this as

$$\sum_r\Lambda^r_{\ell m}\Lambda^r_{\ell' m'}\Delta^2 G^2_r\sigma^2_r\underbrace{\sum_{p=0}^{N_r-1}{\rm e}^{-{\rm i}\phi_p(m-m')}}_{N_r\delta_{mm'}}$$

$$= \Delta\sum_r\Lambda^r_{\ell m}\Lambda^r_{\ell' m}\underbrace{\Delta G^2_r\sigma^2_r}_{G'_r} \equiv h'(\ell, \ell', m). \tag{F12}$$

The sum is equivalent to the previous expression for $h(\ell, \ell', m)$ (equation E5) with a new window G'_r. This motivates the definition of $h(\ell, \ell', m, m')$ such that

$$M^{\rm N}_{\ell\ell'} = \frac{2}{(2\ell+1)(2\ell'+1)}\sum_{mm'}h'^2(\ell, \ell', m, m'), \tag{F13}$$

where

$$h'(\ell, \ell', m, m') \equiv \Delta\sum_j G'_j Y^j_{\ell m}Y^j_{\ell' m'}, \tag{F14}$$

where $G'_j = \Delta G^2_r\sigma^2_{r,p}$. These function can also be calculated using the recursion which we deduce in appendix (C). Note that the noise correlation matrix usually is diagonally dominant and calculating

only the elements close to the diagonal suffices and speeds up the calculations.

One can then find the total correlation matrix, splitting it up into one part due to signal, one part due to noise and a cross term,

$$\tilde{a}_{\ell m} = \tilde{a}^{\rm S}_{\ell m} + \tilde{a}^{\rm N}_{\ell m} \tag{F15}$$

$$\tilde{C}_\ell = \frac{1}{2\ell+1}\sum_m \langle \tilde{a}_{\ell m}\tilde{a}_{\ell' m}\rangle = \tilde{C}^{\rm S}_\ell + \tilde{C}^{\rm N}_\ell + \tilde{C}^{\rm X}_\ell \tag{F16}$$

$$\langle \tilde{C}_\ell \rangle = \left\langle \tilde{C}^{\rm S}_\ell \right\rangle + \left\langle \tilde{C}^{\rm N}_\ell \right\rangle \tag{F17}$$

$$\tilde{C}^{\rm X}_\ell = \frac{1}{2\ell+1}\sum_m \left(a^{\rm S}_{\ell m} a^{\rm N*}_{\ell m} + a^{\rm N}_{\ell m} a^{\rm S*}_{\ell m}\right), \tag{F18}$$

where the assumption that there is no correlation between signal and noise was used. One can then see that the correlation matrix can be written in a similar manner,

$$\left\langle \tilde{C}_{\ell_i}\tilde{C}_{\ell_j}\right\rangle - \left\langle \tilde{C}_{\ell_i}\right\rangle\left\langle \tilde{C}_{\ell_j}\right\rangle = M^{\rm S}_{ij} + M^{\rm N}_{ij} + \left\langle \tilde{C}^{\rm X}_{\ell_i}\tilde{C}^{\rm X}_{\ell_j}\right\rangle. \tag{F19}$$

This is another major result of this paper showing the full correlation matrix of $\tilde{C}_\ell$ including noise. One can write the cross term as,

$$\left\langle \tilde{C}^{\rm X}_\ell \tilde{C}^{\rm X}_{\ell'}\right\rangle = 4\sum_{mm'}\frac{\left\langle \tilde{a}^{\rm S}_{\ell m}\tilde{a}^{\rm S*}_{\ell' m'}\right\rangle\left\langle \tilde{a}^{\rm N}_{\ell m}\tilde{a}^{\rm N*}_{\ell' m'}\right\rangle}{(2\ell+1)(2\ell'+1)} \tag{F20}$$

$$= 4\sum_{m}\frac{\left\langle \tilde{a}^{\rm S}_{\ell m}\tilde{a}^{\rm S*}_{\ell' m}\right\rangle\left\langle \tilde{a}^{\rm N}_{\ell m}\tilde{a}^{\rm N*}_{\ell' m}\right\rangle}{(2\ell+1)(2\ell'+1)}, \tag{F21}$$

where the relation $\langle \tilde{a}^{\rm S}_{\ell m}\tilde{a}^{\rm S*}_{\ell' m'}\rangle = \delta_{mm'}\langle \tilde{a}^{\rm S}_{\ell m}\tilde{a}^{\rm S*}_{\ell' m}\rangle$ was used. From the above, one can see that these two factors can be written as,

$$\langle \tilde{a}^{\rm S}_{\ell m}\tilde{a}^{\rm S*}_{\ell' m}\rangle = \sum_{\ell''} C_{\ell''} h(\ell,\ell'',m)h(\ell',\ell'',m), \tag{F22}$$

$$\langle \tilde{a}^{\rm N}_{\ell m}\tilde{a}^{\rm N*}_{\ell' m}\rangle = \sum_i G_i^2 Y^i_{\ell m} Y^i_{\ell' m}\Delta_i^2\sigma_i^2, \tag{F23}$$

$$= h'(\ell,\ell',m). \tag{F24}$$

Again using the binning in equation (24), one can write the signal-noise cross correlation matrix similar to equation (E9) as

$$M^{\rm X}_{ij} \equiv \langle \tilde{C}^{\rm X}_{\ell_i}\tilde{C}^{\rm X}_{\ell_j}\rangle = \sum_k D_b \chi'(b,i,j), \tag{F25}$$

where

$$\chi'(b,i,j) \equiv \frac{2}{(2\ell_i+1)(2\ell_j+1)} \times \sum_m \left[\sum_{l\epsilon b} B_\ell^2 \ell(\ell+1)h(\ell,\ell_i,m)h(\ell,\ell_j,m)\right]\cdot \times h'(\ell_i,\ell_j,m) \tag{F26}$$

APPENDIX G: DERIVATIVES OF THE LIKELIHOOD

In the minimization of the likelihood, one also needs the first and second derivative of the log-likelihood with respect to the bin values D_b described in equation (22). These can be found to be

$$\frac{\partial L}{\partial D_b} = Tr(\mathbf{A}_b) + \boldsymbol{f}\boldsymbol{d} \tag{G1}$$

$$\frac{\partial^2 L}{\partial D_b \partial D_{b'}} = -Tr(\mathbf{A}_b\mathbf{A}_{b'}) + Tr\left(\mathbf{C}^{-1}\frac{\partial^2\mathbf{C}}{\partial D_b\partial D_{b'}}\right) + 2\boldsymbol{h}_b^{\rm T}\mathbf{C}^{-1}\boldsymbol{h}_{b'} \tag{G2}$$

$$- 2\boldsymbol{h}_b^{\rm T}\boldsymbol{g}_{b'} - 2\boldsymbol{h}_{b'}^{\rm T}\boldsymbol{g}_b - \boldsymbol{f}^{\rm T}\frac{\partial^2\mathbf{C}}{\partial D_b\partial D_{b'}}\boldsymbol{f} + 2\boldsymbol{f}^{\rm T}\boldsymbol{k}_{bb'} + 2\frac{\partial \boldsymbol{d}^{\rm T}}{\partial D_b}\boldsymbol{g}_{b'} \tag{G3}$$

We have used the following definitions,

$$\mathbf{A}_b = \mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial D_b} \tag{G4}$$

$$\boldsymbol{h}_b = \frac{\partial\mathbf{C}}{\partial D_b}\mathbf{C}^{-1}\boldsymbol{d} \tag{G5}$$

$$\boldsymbol{g}_b = \mathbf{C}^{-1}\frac{\partial\boldsymbol{d}}{\partial D_b} \tag{G6}$$

$$\boldsymbol{f} = \mathbf{C}^{-1}\boldsymbol{d} \tag{G7}$$

$$\boldsymbol{k}_{bb'} = \frac{\partial^2\boldsymbol{d}}{\partial D_b\partial D_{b'}} \tag{G8}$$

Here the derivatives of $\boldsymbol{d}$ are,

$$\frac{\partial d_i}{\partial D_b} = -\frac{\partial\langle\tilde{C_{\ell_i}}\rangle}{\partial D_b} = \frac{-1}{2\ell_i+1}\sum_m\left[\sum_{L'} h(\ell_i,L',m)^2\frac{\partial C_{L'}}{\partial D_b}B_\ell^2\right], \tag{G9}$$

$$\frac{\partial^2 d_i}{\partial D_b\partial D_{b'}} = -\frac{\partial^2\langle\tilde{C_{\ell_i}}\rangle}{\partial D_b\partial D_{b'}} = \frac{-1}{2\ell_i+1}\sum_m \times\left[\textstyle\sum_{L'} h(\ell_i,L',m)^2\frac{\partial^2 C_{L'}}{\partial D_b\partial D_{b'}}B_\ell^2\right]. \tag{G10}$$

Obviously for our binning, the double derivative of $\boldsymbol{d}$ disappears.

APPENDIX H: CORRELATION BETWEEN DIFFERENT PATCHES

Suppose one has two axisymmetric Gabor windows, $G^{\rm A}(\hat{\boldsymbol{n}})$ and $G^{\rm B}(\hat{\boldsymbol{n}})$, centred at two different positions A and B on the sky. Suppose also that the rotation operators $\hat{D}^{\rm A}$ and $\hat{D}^{\rm B}$ will rotate these patches so that the centres are on the north pole. Considering patch A, one can define,

$$\tilde{a}^{\rm A}_{\ell m} = \int G^{\rm A}_0(\hat{\boldsymbol{n}})[\hat{D}^{\rm A}T(\hat{\boldsymbol{n}})]Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{H1}$$

where $G_0^{\rm A}$ is the window $G^{\rm A}$ rotated to the north pole. As $T(\hat{\boldsymbol{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}})$, one gets that

$$\hat{D}^{\rm A} T(\hat{\boldsymbol{n}}) = \sum_{\ell m} a_{\ell m} \sum_{m'} D^{lA}_{m'm} Y_{\ell m'}(\hat{\boldsymbol{n}}). \tag{H2}$$

Here the $D^{\ell}_{m'm}$ coefficients are described in Appendix A. One now gets,

$$\tilde{a}^{\rm A}_{\ell m} = \sum_{\ell' m'} a_{\ell' m'} \sum_{m''} D^{\ell' A}_{m'' m'} h^{\rm A}(\ell, \ell', m) \delta_{m m''} \tag{H3}$$

$$= \sum_{\ell' m'} a_{\ell' m'} D^{\ell' A}_{m m'} h^{\rm A}(\ell, \ell', m), \tag{H4}$$

where $h^{\rm A}(\ell, \ell', m)$ is just the $h(\ell, \ell', m)$ function for the Gabor window $G^{\rm A}(\hat{\boldsymbol{n}})$.

The next step is to find the correlations between $\tilde{C}^{\rm A}_{\ell}$ and $\tilde{C}^{\rm B}_{\ell}$, defined for patch A as,

$$\tilde{C}^{\rm A}_{\ell} = \sum_{m} \frac{\tilde{a}^{\rm A}_{\ell m} \tilde{a}^{\rm A*}_{\ell m}}{2\ell + 1}. \tag{H5}$$

Following the procedure we used for a single patch one gets,

$$\langle \tilde{C}^{\rm A}_{\ell} \tilde{C}^{\rm B}_{\ell'} \rangle = \frac{1}{(2\ell+1)(2\ell'+1)} \sum_{m m'} \left\langle \tilde{a}^{\rm A*}_{\ell m} \tilde{a}^{\rm A}_{\ell m} \tilde{a}^{\rm B*}_{\ell' m'} \tilde{a}^{\rm B}_{\ell' m'} \right\rangle$$

$$= \left\langle \tilde{C}^{\rm A}_{\ell} \right\rangle \left\langle \tilde{C}^{\rm B}_{\ell'} \right\rangle + \frac{2 \sum_{m m'} |\left\langle \tilde{a}^{\rm A*}_{\ell m} \tilde{a}^{\rm B}_{\ell' m'} \right\rangle|^2}{(2\ell+1)(2\ell'+1)}. \tag{H6}$$

One can use the expression for $\tilde{a}^{\rm A}_{\ell m}$ to find,

$$\langle \tilde{a}^{\rm A}_{\ell m} \tilde{a}^{\rm B*}_{\ell m} \rangle = \sum_{\ell'' m''} \sum_{L'' M''} \langle a_{\ell'' m''} a^*_{L'' M''} \rangle D^{\ell'' A}_{m m''} D^{L'' B*}_{m' M''} \tag{H7}$$

$$\times h^{\rm A}(\ell, \ell'', m) h^{\rm B}(\ell', L'', m') \tag{H8}$$

$$= \sum_{\ell''} C_{\ell''} \underbrace{\left(\sum_{m''} D^{\ell'' A}_{m m''} D^{\ell'' B*}_{m' m''} \right)}_{D^{\ell''}_{m m'}(\Delta)} h^{\rm A}(\ell, \ell'', m) h^{\rm B}(\ell', \ell'', m') \tag{H9}$$

$$= \sum_{\ell''} C_{\ell''} d^{\ell''}_{m m'}(\Delta) h^{\rm A}(\ell, \ell', m) h^{\rm B}(\ell', \ell'', m'), \tag{H10}$$

where Δ is the angel between the centres of the patches. Relations from Appendix A were used here.

The next step is to see what happens when noise is introduced. We assume that the noise is uncorrelated. The noise in pixel j is n_j and $\langle n_j n'_j \rangle = \delta_{jj'} \sigma_j^2$. From above one has,

$$\tilde{C}^{\rm A}_{\ell} = \tilde{C}^{\rm AS}_{\ell} + \tilde{C}^{\rm AN}_{\ell} + \tilde{C}^{\rm AX}_{\ell}, \tag{H11}$$

where

$$\tilde{C}^{\rm AN}_{\ell} = \sum_{m} \frac{\tilde{a}^{\rm NA}_{\ell m} \tilde{a}^{\rm NA*}_{\ell m}}{2\ell + 1} \tag{H12}$$

$$\tilde{C}^{\rm AX}_{\ell} = \sum_{m} \frac{\tilde{a}^{\rm NA}_{\ell m} \tilde{a}^{\rm SA*}_{\ell m}}{2\ell + 1} \tag{H13}$$

$$\tilde{a}^{\rm AN}_{\ell m} = \sum_{j} G^{\rm A}_j n^{\rm A}_j Y^j_{\ell m}, \tag{H14}$$

where the last sum is over pixels, $G^{\rm A}_j$ and $n^{\rm A}_j$ being the window and noise for pixel j respectively.

The correlation between the two patches then becomes,

$$\left\langle \tilde{C}^{\rm A}_{\ell} \tilde{C}^{\rm B}_{\ell'} \right\rangle - \left\langle \tilde{C}^{\rm A}_{\ell} \right\rangle \left\langle \tilde{C}^{\rm B}_{\ell'} \right\rangle = M^{\rm S}_{\ell\ell'} + M^{\rm N}_{\ell\ell'} + \left\langle \tilde{C}^{\rm AX}_{\ell} \tilde{C}^{\rm BX}_{\ell'} \right\rangle, \tag{H15}$$

where,

$$M^{\rm S}_{\ell\ell'} = \frac{2}{(2\ell+1)(2\ell'+1)} \sum_{m m'} \left| \left\langle \tilde{a}^{\rm AS*}_{\ell m} \tilde{a}^{\rm BS}_{\ell' m'} \right\rangle \right|^2 \tag{H16}$$

$$M^{\rm N}_{\ell\ell'} = \frac{2}{(2\ell+1)(2\ell'+1)} \sum_{m m'} \left| \left\langle \tilde{a}^{\rm AN*}_{\ell m} \tilde{a}^{\rm BN}_{\ell' m'} \right\rangle \right|^2. \tag{H17}$$

Finally,

$$\left\langle \tilde{C}^{\rm AX}_{\ell} \tilde{C}^{\rm BX}_{\ell'} \right\rangle = \frac{4}{(2\ell+1)(2\ell'+1)} \sum_{m m'} \left\langle \tilde{a}^{\rm AS*}_{\ell m} \tilde{a}^{\rm BS}_{\ell' m'} \right\rangle \left\langle \tilde{a}^{\rm AN*}_{\ell m} \tilde{a}^{\rm BN}_{\ell' m'} \right\rangle. \tag{H18}$$

Now one needs an expression for $\langle \tilde{a}^{\rm AN}_{\ell m} \tilde{a}^{\rm BN*}_{\ell' m'} \rangle$. One gets,

$$\left\langle \tilde{a}^{\rm AN}_{\ell m} \tilde{a}^{\rm BN*}_{\ell' m'} \right\rangle = \sum_{j j'} G^{\rm A}_j G^{\rm B}_{j'} \left\langle n^{\rm A}_j n^{\rm B}_{j'} \right\rangle Y^j_{\ell m} Y^{j'}_{\ell' m'}. \tag{H19}$$

Here there are only correlations between overlapping pixels. If there are no overlapping pixels between the patches, this term is zero. Otherwise this can be written as a sum over the overlapping pixels

$$\langle \tilde{a}^{\rm AN}_{\ell m} \tilde{a}^{\rm BN*}_{\ell' m'} \rangle = \sum_{j} G^{\rm A}_j G^{\rm B}_j \sigma^2_j Y^j_{\ell m} Y^{j'}_{\ell' m'}. \tag{H20}$$

This paper has been typeset from a TEX/LATEX file prepared by the author.